

© 2017 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Joint Sensor and Relay Power Control in Tracking Gaussian Mixture Targets by Wireless Sensor Networks

Johann A. Bengua, Hoang Duong Tuan, Trung Q. Duong and H. Vincent Poor

Abstract—This paper considers a wireless sensor network (WSN) for locating a static target or tracking a dynamic target, which is prior characterized by a Gaussian mixture (GM) distribution. An amplify-and-forward relay node is acting as a wireless bridge in relaying the sensor’s independent observations of the target to a fusion center (FC). We are concerned with the joint transmitter power allocation for the sensors and relay to optimize the Bayesian filter, which is deployed at the FC for a global estimate of the target. The mean squared error (MSE) of the Bayesian filter is already computationally intractable for fixed sensor and relay transmitter power, so power allocation to minimize its mean squared error is a very challenging problem. In this paper, the problem is addressed by an iterative procedure of very low computational complexity. Simulations are provided to support the efficiency of our proposed power allocation.

Index Terms—Bayesian filter, Gaussian mixture target, target localization, target tracking, wireless sensor networks, relaying communication, data fusion, fast computation

I. INTRODUCTION

Wireless sensor networks (WSNs), which consist of spatially distributed wireless sensors, play a key role in many applications such as process monitoring in industrial plants, navigational and guidance systems, radar tracking, sonar ranging, environment monitoring, battlefield surveillance, health care and Internet of Things (IoT) [1]–[12]. Each sensor in the network often operates in an amplify-and-forward mode [13], [14] in delivering its local observation on a target to a central system, known as the fusion center (FC). The FC filters these local observations for a global estimate of the target. The sensors may be linear or nonlinear depending upon their input-output relations. For instance, the ranging and/or bearing sensors [4] for target localization and tracking are nonlinear. The target is often assumed to be prior Gaussian in linear

models, in which case the Bayesian filter is defined via the first and second order statistical moments of the jointly Gaussian distributed source and observation [15, p. 155]. As the sensors are limited by energy resources, sensor transmitter power allocation in linear sensor networks (LSNs) via minimizing estimate distortion at the FC for scalar Gaussian targets has been a subject of considerable interest [16]–[21]. Provided that the target is prior characterized by a Gaussian random variable, our previous work [22] derived a tractable semi-definite program (SDP) for sensor power allocation in both LSNs and nonlinear sensor networks (NSNs). The SDP allows the FC to determine the best linear estimate in terms of the mean squared error (MSE) irrespective of targets that are scalar or vector, static or dynamic. The wireless communication channels between the sensors and FC have been assumed strong enough in all aforementioned works to compensate the sensors’ low transmitter power. As all wireless channels suffer the common impairments such as path-loss, shadowing and small-scale fading, this assumption implicitly implies that the sensors must be in a good position relative to the FC, which is not always possible. It is known that wireless relay nodes can be deployed to act as wireless bridges to effectively assist the communication between the sensors and the FC. Multi-hop communication/relaying has been accepted as a standard to provide high capacity coverage area in next generation wireless broadband systems [23]. However, the relaying techniques for wireless transmission have not been explored in wireless sensor networks.

Meanwhile, Gaussian mixture models (GMMs) have been widely acknowledged as a better means than Gaussian models in characterizing the target priors since they offer more useful information [24]. Indeed, GMMs have been shown to provide powerful tools in signal processing (see e.g. [25]–[27] and references therein). However, Bayesian filters for Gaussian mixture targets¹ already causes computational intractability in linear models, simply because there is no closed-form of the MSE function. A particular problem has been addressed in [27] by stochastic programming.

In this paper we will address the joint sensor and relay power allocation to optimize Bayesian filters in estimating static or dynamic targets with Gaussian mixture prior knowledge by LSNs or NSNs, which non-trivially changes the nature of the power allocation and requires a different approach to the

This work was supported in part by the Australian Research Council’s Discovery Projects under Project DP130104617, in part by the U.K. Royal Academy of Engineering Research Fellowship under Grant RF1415/14/22, in part by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/P019374/1, and in part by the U.S. National Science Foundation under Grants CNS-145679 and CNS-1702808.

Johann A. Bengua was with the Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia. He is now with Teradata Australia and New Zealand (Email: jbengua@gmail.com).

Hoang D. Tuan is with the School of Electrical and Data Engineering, University of Technology Sydney, Ultimo, NSW 2007, Australia (Email: tuan.hoang@uts.edu.au)

Trung Q. Duong is with Queen’s University Belfast, Belfast BT7 1NN, U.K. (Email: trung.q.duong@qub.ac.uk)

H. Vincent Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

¹For convenience of presentation, Gaussian mixture targets are referred to those that are prior Gaussian mixtures.

solution. To the authors's best knowledge, this problem has not been considered in the literature. Our contribution is to show that this computationally intractable problem can be addressed by an iterative scalable procedure of very low computational complexity, which converges to a stationary point after only a few iterations.

The rest of the paper is structured as follows. Section II is devoted to a general sensor and relay power allocation to optimize the Bayes filter and its computational procedure. Its applications to stationary Gaussian mixture target localization by LSNs and NSNs are presented in Section III. Section IV presents the applications to Kalman filtering for tracking a dynamic Gaussian mixture target by LSNs and NSNs. Section V concludes the paper. To make the paper self-contained we also provide in Appendix A some fundamental facts on Gaussian mixture with new matrix inequalities, which play an important role in our algorithmic developments.

Notation. Bold lower-case and upper-case symbols are used to represent vectors and matrices, respectively. By $\mathbf{A} \succeq \mathbf{B}$ it means $\mathbf{A} - \mathbf{B} \succeq \mathbf{0}$, i.e. $\mathbf{A} - \mathbf{B}$ is a positive semi-definite matrix. $\mathbf{x} > \mathbf{0}$ for a vector $\mathbf{x} = (x_1, \dots, x_m)^T$ is component-wise understood, i.e. $x_i > 0$, $i = 1, \dots, m$, while $(\mathbf{x})^2$ is the positive semi-definite matrix $\mathbf{x}\mathbf{x}^T$. \mathbf{I}_n is the identity matrix of size $n \times n$. $\text{diag}[a_i]_1^N$ or $\text{diag}[a_i]_{i=1, \dots, N}$ or $\text{diag}[a_1, a_2, \dots, a_N]$ is a diagonal matrix with ordered diagonal entries a_1, a_2, \dots, a_N , which may be scalars or matrices. The trace of a square matrix \mathbf{A} is expressed by $\text{Trace}(\mathbf{A})$. $\mathbb{E}[\cdot]$ is the expectation operator. $\mathbf{X} \sim p_{\mathbf{X}}(\cdot)$ is referred to a random variable (RV) \mathbf{X} with probability density function (PDF) $p_{\mathbf{X}}(\cdot)$. $\mathbf{m}_{\mathbf{X}}$ is its expectation $\mathbb{E}[\mathbf{X}]$, while $\mathbf{C}_{\mathbf{X}}$ is its auto-covariance matrix $\mathbb{E}[(\mathbf{X} - \mathbf{m}_{\mathbf{X}})^2]$ and $\mathbf{C}_{\mathbf{X}\mathbf{Y}}$ is its cross-covariance matrix $\mathbb{E}[(\mathbf{X} - \mathbf{m}_{\mathbf{X}})(\mathbf{Y} - \mathbf{m}_{\mathbf{Y}})^T]$ with another RV \mathbf{Y} . Similarly $\mathbf{R}_{\mathbf{X}}$ is its auto-correlation matrix $\mathbb{E}[\mathbf{X}\mathbf{X}^T] = \mathbf{C}_{\mathbf{X}} + (\mathbf{m}_{\mathbf{X}})^2$ and $\mathbf{R}_{\mathbf{X}\mathbf{Y}}$ is its cross-correlation matrix $\mathbb{E}[\mathbf{X}\mathbf{Y}^T] = \mathbf{C}_{\mathbf{X}\mathbf{Y}} + \mathbf{m}_{\mathbf{X}}(\mathbf{m}_{\mathbf{Y}})^T$ with another RV \mathbf{Y} . $\mathbf{X}|\mathbf{Y}$ is a random variable \mathbf{X} restricted by a realization of the conditioning random variable \mathbf{Y} and accordingly $\mathbf{X}|\mathbf{Y} = \mathbf{y}$ is a random variable restricted by the value $\mathbf{Y} = \mathbf{y}$ of \mathbf{Y} . $\mathcal{N}(\mathbf{x}; \mathbf{m}_{\mathbf{X}}, \mathbf{C}_{\mathbf{X}}) := \frac{1}{\sqrt{2\pi \det(\mathbf{C}_{\mathbf{X}})}} \exp(-\frac{1}{2}(\mathbf{x} - \mathbf{m}_{\mathbf{X}})^T \mathbf{C}_{\mathbf{X}}^{-1}(\mathbf{x} - \mathbf{m}_{\mathbf{X}}))$ is a Gaussian distribution so $\mathbf{X} \sim \mathcal{N}(\cdot; \mathbf{m}_{\mathbf{X}}, \mathbf{C}_{\mathbf{X}})$ means that \mathbf{X} is Gaussian random variable (RV) with expectation $\mathbf{m}_{\mathbf{X}}$ and covariance $\mathbf{C}_{\mathbf{X}}$.

II. JOINT GMM RELAYED EQUATIONS

In statistical signal processing, detection and estimation of an object is based on the knowledge of its statistics along with noisy observations [15]. We start the section by introducing the following joint GMM for the N -dimensional target \mathbf{X} and M -sensor noisy observation \mathbf{Y}

$$(\mathbf{X}, \mathbf{Y}) \sim \sum_{i=1}^L \lambda_i \mathcal{N}\left((\cdot, \cdot); \mathbf{m}_{\mathbf{X}, \mathbf{Y}}^{(i)}, \mathbf{C}^{(i)}\right) \quad (1)$$

with $\lambda_i > 0$, $\sum_{i=1}^L \lambda_i = 1$ and $\mathbf{m}_{\mathbf{X}, \mathbf{Y}}^{(i)} = \begin{pmatrix} \mathbf{m}_{\mathbf{X}}^{(i)} \\ \mathbf{m}_{\mathbf{Y}}^{(i)} \end{pmatrix}$,

$$\mathbf{C}^{(i)} = \begin{pmatrix} \mathbf{C}_{\mathbf{X}}^{(i)} & \mathbf{C}_{\mathbf{X}\mathbf{Y}}^{(i)} \\ \mathbf{C}_{\mathbf{Y}\mathbf{X}}^{(i)} & \mathbf{C}_{\mathbf{Y}}^{(i)} \end{pmatrix}. \text{ It is well known (see e.g. [15,$$

Chapter III], [22], [28]) that almost all results for Gaussian target estimation are based on the derivation of the joint Gaussian distribution of the target and its observation. We will see later in the present paper that the joint GM distribution (1) facilitates unified framework for Bayesian and Kalman filters in both linear and nonlinear models. For the reader's convenience, we provide in Appendix A the minimum background on GMMs with some new results, which motivates our development in this paper.

In accordance to GMM (1), \mathbf{X} is a Gaussian mixture (GM) $\sum_{i=1}^L \lambda_i \mathcal{N}(\mathbf{x}; \mathbf{m}_{\mathbf{X}}^{(i)}, \mathbf{C}_{\mathbf{X}}^{(i)})$ and $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_M)^T$ is a GM $\sum_{i=1}^L \lambda_i \mathcal{N}(\mathbf{y}; \mathbf{m}_{\mathbf{Y}}^{(i)}, \mathbf{C}_{\mathbf{Y}}^{(i)})$. The sensor observations are instantaneously sampled value $\mathbf{y} = (y_1, y_2, \dots, y_M)^T$ of \mathbf{Y} . One can define

$$\mathbb{E}[|y_j|^2] = \mathbf{C}_{\mathbf{Y}}(j, j) + \mathbf{m}_{\mathbf{Y}}^2(j), \quad (2)$$

where $\mathbf{C}_{\mathbf{Y}}$ is the covariance matrix of \mathbf{Y} and $\mathbf{m}_{\mathbf{Y}} = \mathbb{E}[\mathbf{Y}]$. As illustrated by Figure 1, the sensors send these observations y_j to the relay over wireless time-orthogonal communication channels [16]. The analog signals received at the relay can thus be written as

$$z_{jR} = \sqrt{h_{jR}} \alpha_j y_j + w_{jR}, \quad j = 1, 2, \dots, M, \quad (3)$$

where $\sqrt{h_{jR}}$ is the channel gain between sensor j and the relay, w_{jR} is a corrupt noise, which can be assumed white with power σ_{jR} and independent from y_j , and $\sqrt{\alpha_j}$ controls the transmitter power $P_j = \alpha_j \mathbb{E}[|y_j|^2] = (\mathbf{C}_{\mathbf{Y}}(j, j) + \mathbf{m}_{\mathbf{Y}}^2(j)) \alpha_j$ of sensor j , which is subject to (s.t.) a fixed sum power budget $P_T > 0$, defined as

$$\sum_{j=1}^M P_j = \sum_{j=1}^M \mathbb{E}[|y_j|^2] \alpha_j \leq P_T. \quad (4)$$

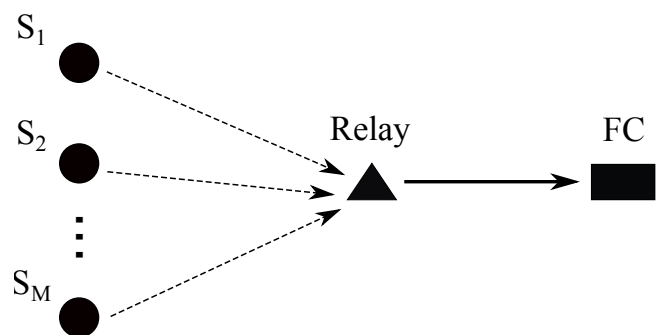


Fig. 1. System model

According to Figure 1, the relay will then amplify these received signals z_{jR} to power level β_j before forwarding them to the FC over wireless time-orthogonal communication channels so the analog signals received at the FC are

$$\begin{aligned} z_j &= \sqrt{h_{jD}} \sqrt{\beta_j / |z_{jR}|^2} z_{jR} + w_{jD} \\ &= \sqrt{h_{jD} h_{jR} \beta_j \alpha_j / (h_{jR} \mathbb{E}[|y_j|^2] \alpha_j + \sigma_{jR})} y_j + w_{jD}, \end{aligned} \quad (5)$$

where $\sqrt{h_{jD}}$ is the channel gain between the relay and the FC, w_{jD} is the background noise at the FC, which can be

assumed to be noise with power σ_{jD} and independent from z_{jR} . Accordingly,

$$w_j = \sqrt{h_{jD}\beta_j/(h_{jR}\mathbb{E}[|y_j|^2]\alpha_j + \sigma_{jR})}w_{jR} + w_{jD}$$

is white noise with power $\sigma_{jR}h_{jD}\beta_j/(h_{jR}\mathbb{E}[|y_j|^2]\alpha_j + \sigma_{jR}) + \sigma_{jD}$. The power levels β_j are constrained by the relay power budget P_R , defined as

$$\sum_{i=1}^M \beta_j \leq P_R. \quad (6)$$

Thus, the signals received at the FC can be written in a vector form by

$$\mathbf{Z} = \mathbf{H}_{\alpha,\beta}\mathbf{Y} + \mathbf{W}_{\alpha,\beta}, \quad (7)$$

where $\mathbf{H}_{\alpha,\beta} \in \mathbb{R}^{M \times M}$ is defined by

$$\mathbf{H}_{\alpha,\beta} = \text{diag} \left[\sqrt{h_{jD}h_{jR}\beta_j\alpha_j/(h_{jR}\mathbb{E}[|y_j|^2]\alpha_j + \sigma_{jR})} \right]_1^M,$$

and the total noise $\mathbf{W}_{\alpha,\beta} \sim \mathcal{N}(\cdot; 0, \mathbf{C}_{\alpha,\beta})$ with diagonal matrix

$$\mathbf{C}_{\alpha,\beta} = \text{diag}[\sigma_{jR}h_{jD}\beta_j/(h_{jR}\mathbb{E}[|y_j|^2]\alpha_j + \sigma_{jR}) + \sigma_{jD}]_1^M.$$

Based on the joint GM (1) for the target \mathbf{X} and sensor noisy observation \mathbf{Y} and the output equation (7) for relayed observation \mathbf{Z} , one can write the joint distribution of the target \mathbf{X} and its relayed observations \mathbf{Z} at FC as

$$(\mathbf{X}, \mathbf{Z}) \sim \sum_{i=1}^L \lambda_i \mathcal{N} \left(\left(\cdot, \cdot \right); \begin{pmatrix} \mathbf{m}_{\mathbf{X}}^{(i)} \\ \mathbf{H}_{\alpha,\beta} \mathbf{m}_{\mathbf{Y}}^{(i)} \end{pmatrix}, \begin{pmatrix} \mathbf{C}_{\mathbf{X}}^{(i)} & \mathbf{C}_{\mathbf{X}\mathbf{Y}}^{(i)} \mathbf{H}_{\alpha,\beta} \\ \mathbf{H}_{\alpha,\beta} \mathbf{C}_{\mathbf{Y}\mathbf{X}}^{(i)} & \mathbf{H}_{\alpha,\beta} \mathbf{C}_{\mathbf{Y}}^{(i)} \mathbf{H}_{\alpha,\beta} + \mathbf{C}_{\alpha,\beta} \end{pmatrix} \right).$$

Accordingly,

$$\{\mathbf{X}|\mathbf{Z} = \mathbf{z}\} \sim \sum_{i=1}^L \lambda_i(\mathbf{z}, \alpha, \beta) \mathcal{N}(\cdot, \mathbf{m}_{\mathbf{X}^{(i)}|\mathbf{Z}^{(i)}}), \quad (8)$$

where

$$\mathbf{m}_{\mathbf{X}^{(i)}|\mathbf{Z}^{(i)}} = \mathbf{m}_{\mathbf{X}}^{(i)} + \mathbf{C}_{\mathbf{Y}\mathbf{X}}^{(i)T} \mathbf{H}_{\alpha,\beta} (\mathbf{H}_{\alpha,\beta} \mathbf{C}_{\mathbf{Y}}^{(i)} \mathbf{H}_{\alpha,\beta} + \mathbf{C}_{\alpha,\beta})^{-1} (\mathbf{z} - \mathbf{H}_{\alpha,\beta} \mathbf{m}_{\mathbf{Y}}^{(i)}), \quad (9)$$

$$\lambda_i(\mathbf{z}, \alpha, \beta) = \frac{\lambda_i \mathcal{N}(\mathbf{z}; \mathbf{H}_{\alpha,\beta} \mathbf{m}_{\mathbf{Y}}^{(i)}, \mathbf{H}_{\alpha,\beta} \mathbf{C}_{\mathbf{Y}}^{(i)} \mathbf{H}_{\alpha,\beta} + \mathbf{C}_{\alpha,\beta})}{\sum_{i=1}^L \lambda_i \mathcal{N}(\mathbf{z}; \mathbf{H}_{\alpha,\beta} \mathbf{m}_{\mathbf{Y}}^{(i)}, \mathbf{H}_{\alpha,\beta} \mathbf{C}_{\mathbf{Y}}^{(i)} \mathbf{H}_{\alpha,\beta} + \mathbf{C}_{\alpha,\beta})}, \quad (10)$$

$$\mathbf{C}_{\mathbf{X}^{(i)}|\mathbf{Z}^{(i)}} = \mathbf{C}_{\mathbf{X}}^{(i)} - \mathbf{C}_{\mathbf{Y}\mathbf{X}}^{(i)T} \mathbf{H}_{\alpha,\beta} (\mathbf{H}_{\alpha,\beta} \mathbf{C}_{\mathbf{Y}}^{(i)} \mathbf{H}_{\alpha,\beta} + \mathbf{C}_{\alpha,\beta})^{-1} \mathbf{H}_{\alpha,\beta} \mathbf{C}_{\mathbf{Y}\mathbf{X}}^{(i)}. \quad (11)$$

The Bayesian estimate $\hat{\mathbf{x}}(\mathbf{z})$ based on FC output $\mathbf{Z} = \mathbf{z}$ is

$$\hat{\mathbf{x}}(\mathbf{z}) \triangleq \mathbb{E}[\mathbf{X}|\mathbf{Z} = \mathbf{z}] = \sum_{i=1}^L \lambda_i(\mathbf{z}, \alpha, \beta) \mathbf{m}_{\mathbf{X}^{(i)}|\mathbf{Z}^{(i)}} \quad (12)$$

with the mean squared error

$$\mathbb{E}(\|\hat{\mathbf{x}}(\mathbf{z}) - \mathbf{x}\|^2) = \text{Trace}(\mathbf{C}_{\mathbf{z}}(\alpha, \beta)), \quad (13)$$

where

$$\mathbf{C}_{\mathbf{z}}(\alpha, \beta) = \sum_{i=1}^L \lambda_i(\mathbf{z}, \alpha, \beta) [\mathbf{C}_{\mathbf{X}^{(i)}|\mathbf{Z}^{(i)}} + (\mathbf{m}_{\mathbf{X}^{(i)}|\mathbf{Z}^{(i)}})^2] - \left(\sum_{i=1}^L \lambda_i(\mathbf{z}, \alpha, \beta) \mathbf{m}_{\mathbf{X}^{(i)}|\mathbf{Z}^{(i)}} \right)^2 \quad (14)$$

(see (66)-(67) in the Appendix A). By defining

$$g(\alpha, \beta) = \mathbb{E}_{\mathbf{z}}(\mathbb{E}(\|\hat{\mathbf{x}}(\mathbf{z}) - \mathbf{x}\|^2)) = \mathbb{E}_{\mathbf{z}}(\text{Trace}(\mathbf{C}_{\mathbf{z}}(\alpha, \beta))),$$

where $\mathbb{E}_{\mathbf{z}}$ is the expectation with respect to random variable \mathbf{z} , the joint sensor and relay power allocation to minimize the mean squared error of Bayesian filtering is formulated by

$$\min_{\alpha, \beta} g(\alpha, \beta) \quad \text{s.t.} \quad (4), (6). \quad (15)$$

Unfortunately, there is no closed-form for function $g(\alpha, \beta)$, making the optimization (15) computationally intractable.

We now use a surrogate function for $g(\alpha, \beta)$ such that the optimization for the former leads to the optimization for the latter. It follows from Theorem 3 in the Appendix A that

$$\mathbf{C}_{\mathbf{z}}(\alpha, \beta) \preceq \mathbf{C}_{lmse}(\alpha, \beta) \quad \forall \mathbf{z}, \quad (16)$$

where

$$\begin{aligned} \mathbf{C}_{lmse}(\alpha, \beta) &= \\ \mathbf{C}_{\mathbf{X}} - \mathbf{C}_{\mathbf{Y}\mathbf{X}}^T \mathbf{H}_{\alpha,\beta} (\mathbf{H}_{\alpha,\beta} \mathbf{C}_{\mathbf{Y}} \mathbf{H}_{\alpha,\beta} + \mathbf{C}_{\alpha,\beta})^{-1} \mathbf{H}_{\alpha,\beta} \mathbf{C}_{\mathbf{Y}\mathbf{X}} &= \\ \mathbf{C}_{\mathbf{X}} - \mathbf{C}_{\mathbf{Y}\mathbf{X}} (\mathbf{C}_{\mathbf{Y}})^{-1} \mathbf{C}_{\mathbf{Y}\mathbf{X}} + \mathbf{C}_{\mathbf{Y}\mathbf{X}}^T (\mathbf{C}_{\mathbf{Y}})^{-1} ((\mathbf{C}_{\mathbf{Y}})^{-1} + & \\ \text{diag}[\varphi_j(\alpha_j, \beta_j)]_1^M)^{-1} (\mathbf{C}_{\mathbf{Y}})^{-1} \mathbf{C}_{\mathbf{Y}\mathbf{X}} & \quad (17) \end{aligned}$$

with $\begin{pmatrix} \mathbf{C}_{\mathbf{X}} & \mathbf{C}_{\mathbf{X}\mathbf{Y}} \\ \mathbf{C}_{\mathbf{Y}\mathbf{X}} & \mathbf{C}_{\mathbf{Y}} \end{pmatrix} = \sum_{i=1}^L \lambda_i (\mathbf{C}^{(i)} + (\mathbf{m}_{\mathbf{X},\mathbf{Y}}^{(i)})^2) - (\mathbf{m}_{\mathbf{X},\mathbf{Y}})^2$, which is the covariance matrix of (\mathbf{X}, \mathbf{Y}) , and $\varphi_j(\alpha_j, \beta_j) = p_j \alpha_j \beta_j / (q_j \alpha_j + r_j \beta_j + \sigma_j)$, $p_j = h_{jR} h_{jD}$, $q_j = h_{jR} \sigma_{jD} \mathbb{E}[|y_j|^2]$, $r_j = \sigma_{jR} h_{jD}$, $\sigma_j = \sigma_{jD} \sigma_{jR}$. In fact, $\text{Trace}(\mathbf{C}_{lmse}(\alpha, \beta))$ is the minimum MSE (MMSE) by linear estimator for \mathbf{X} [28].

Therefore, it is true that $g(\alpha, \beta) \leq \text{Trace}(\mathbf{C}_{lmse}(\alpha, \beta)) \forall \alpha, \beta$ and we seek a suboptimal solution of the computationally intractable optimization problem (15) by solving its following majorant minimization²

$$\min_{\alpha > 0, \beta > 0} \text{Trace}(\mathbf{C}_{lmse}(\alpha, \beta)) \quad \text{s.t.} \quad (4), (6). \quad (18)$$

Being closed-form, the objective function $\text{Trace}(\mathbf{C}_{lmse}(\alpha, \beta))$ is easily computed for every (α, β) . However, its optimization is still difficult and now we develop its computational solution. By (17), $\text{Trace}(\mathbf{C}_{lmse}(\alpha, \beta)) = \text{Trace}(\mathbf{C}_{\mathbf{X}} - \mathbf{C}_{\mathbf{Y}\mathbf{X}}^T (\mathbf{C}_{\mathbf{Y}})^{-1} \mathbf{C}_{\mathbf{Y}\mathbf{X}}) + \text{Trace}(\Psi^T (\Phi + \text{diag}[\varphi_j(\alpha_j, \beta_j)]_1^M)^{-1} \Psi)$, where $\Psi = \mathbf{C}_{\mathbf{Y}}^{-1} \mathbf{C}_{\mathbf{Y}\mathbf{X}}$, $\Phi = \mathbf{C}_{\mathbf{Y}}^{-1}$. Therefore, (18) is equivalent to the following program

$$\min_{\alpha > 0, \beta > 0} \varphi(\alpha, \beta) \quad \text{s.t.} \quad (4), (6), \quad (19)$$

²Note that (18) is computed at the FC, for which the FC needs knowledge on all the parameters such as channel gains $\sqrt{h_{jR}}$ and $\sqrt{h_{jD}}$. These channel gains are obtained separately.

for $\varphi(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \text{Trace}(\boldsymbol{\Psi}^T(\boldsymbol{\Phi} + \text{diag}[\varphi_j(\alpha_j, \beta_j)]_1^M)^{-1}\boldsymbol{\Psi})$. Given $(\boldsymbol{\alpha}^{(\kappa)}, \boldsymbol{\beta}^{(\kappa)})$ feasible for (4), (6) we process the following successive approximations. Define

$$\begin{aligned} \varphi_j^{(\kappa)} &= \varphi_j(\alpha_j^{(\kappa)}, \beta_j^{(\kappa)}), \\ 0 < \boldsymbol{\Theta}^{(\kappa)} &= \text{diag}[\varphi_j^{(\kappa)}]_1^M (\boldsymbol{\Phi} + \text{diag}[\varphi_j^{(\kappa)}]_1^M)^{-1} \boldsymbol{\Psi} \\ &\quad \times \boldsymbol{\Psi}^T (\boldsymbol{\Phi} + \text{diag}[\varphi_j^{(\kappa)}]_1^M)^{-1} \text{diag}[\varphi_j^{(\kappa)}]_1^M, \\ 0 < \rho_j^{(\kappa)} &= \boldsymbol{\Theta}^{(\kappa)}(j, j), \end{aligned} \quad (20)$$

where $\boldsymbol{\Theta}^{(\kappa)}(j, j)$ is the j -th diagonal entry of $\boldsymbol{\Theta}^{(\kappa)}$.

Theorem 1: The following inequalities hold true for all $\boldsymbol{\alpha} > 0$ and $\boldsymbol{\beta} > 0$,

$$\begin{aligned} \varphi(\boldsymbol{\alpha}, \boldsymbol{\beta}) &\leq \varphi(\boldsymbol{\alpha}^{(\kappa)}, \boldsymbol{\beta}^{(\kappa)}) + \\ &\quad \sum_{j=1}^M \rho_j^{(\kappa)} \left(\frac{r_j}{p_j \alpha_j} + \frac{q_j}{p_j \beta_j} + \frac{\sigma_j}{p_j \alpha_j \beta_j} - \frac{1}{\varphi_j^{(\kappa)}} \right) \\ &\leq \varphi^{(\kappa)}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \\ &:= \varphi(\boldsymbol{\alpha}^{(\kappa)}, \boldsymbol{\beta}^{(\kappa)}) + \sum_{j=1}^M \rho_j^{(\kappa)} \left[\frac{r_j}{p_j \alpha_j} + \frac{q_j}{p_j \beta_j} \right. \\ &\quad \left. + \frac{\sigma_j}{2p_j} \left(\frac{\alpha_j^{(\kappa)}}{\beta_j^{(\kappa)} \alpha_j^2} + \frac{\beta_j^{(\kappa)}}{\alpha_j^{(\kappa)} \beta_j^2} \right) - \frac{1}{\varphi_j^{(\kappa)}} \right] \end{aligned} \quad (21)$$

Proof. Define a function $\chi(\boldsymbol{\phi}) = \text{Trace}(\boldsymbol{\Psi}^T(\boldsymbol{\Phi} + \text{diag}[1/\phi_j]_1^M)^{-1}\boldsymbol{\Psi})$, which by the Matrix Inverse Lemma [29] is seen as $\text{Trace}(\boldsymbol{\Psi}^T \text{diag}[\phi_j]_1^M \boldsymbol{\Psi}) - \text{Trace}(\boldsymbol{\Psi}^T (\text{diag}[\phi_j]_1^M (\text{diag}[\phi_j]_1^M + \boldsymbol{\Phi}^{-1})^{-1} \text{diag}[\phi_j]_1^M) \boldsymbol{\Psi})$. The function $\chi(\boldsymbol{\phi})$ is thus concave in $\boldsymbol{\phi} = (\phi_1, \dots, \phi_M)^T > 0$ because the first term is obviously linear while the second term is convex [30, Appendix C]. Therefore, for all $\boldsymbol{\phi} > 0$ and $\boldsymbol{\phi}^{(\kappa)} > 0$, it is true that [31]

$$\begin{aligned} \chi(\boldsymbol{\phi}) &\leq \chi(\boldsymbol{\phi}^{(\kappa)}) + \langle \nabla \chi(\boldsymbol{\phi}^{(\kappa)}), \boldsymbol{\phi} - \boldsymbol{\phi}^{(\kappa)} \rangle \\ &= \chi(\boldsymbol{\phi}^{(\kappa)}) + \text{Trace} \left(\text{diag}[1/\phi_j^{(\kappa)}]_1^M (\boldsymbol{\Phi} \right. \\ &\quad \left. + \text{diag}[1/\phi_j^{(\kappa)}]_1^M)^{-1} \boldsymbol{\Psi} \boldsymbol{\Psi}^T (\boldsymbol{\Phi} + \text{diag}[1/\phi_j^{(\kappa)}]_1^M)^{-1} \right. \\ &\quad \left. \times \text{diag}[1/\phi_j^{(\kappa)}]_1^M (\text{diag}[\phi_j - \phi_j^{(\kappa)}]_1^M) \right). \end{aligned} \quad (23)$$

The inequality (21) is obtained by replacing $\phi_j = 1/\varphi_j(\alpha_j, \beta_j) = r_j/p_j \alpha_j + q_j/p_j \beta_j + \sigma_j/p_j \alpha_j \beta_j$ and $\phi_j^{(\kappa)} = 1/\varphi_j(\alpha_j^{(\kappa)}, \beta_j^{(\kappa)})$ into the above inequality (23).

The inequality (22) follows from the inequality $\frac{1}{\alpha_j \beta_j} \leq$

$$\frac{1}{2} \left(\frac{\alpha_j^{(\kappa)}}{\beta_j^{(\kappa)} \alpha_j^2} + \frac{\beta_j^{(\kappa)}}{\alpha_j^{(\kappa)} \beta_j^2} \right). \quad \square$$

Thus function $\varphi^{(\kappa)}$ is a convex majorant of the highly nonconvex function φ . Accordingly, we consider the following majorant minimization

$$\min_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \varphi^{(\kappa)}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \quad \text{s.t.} \quad (4), (6). \quad (24)$$

Proposition 1: Whenever $(\boldsymbol{\alpha}^{(\kappa)}, \boldsymbol{\beta}^{(\kappa)})$ is feasible for (4), (6), the optimal solution $(\boldsymbol{\alpha}^{(\kappa+1)}, \boldsymbol{\beta}^{(\kappa+1)})$ of the convex program

(24) is a feasible point for the nonconvex program (19), which is better than $(\boldsymbol{\alpha}^{(\kappa)}, \boldsymbol{\beta}^{(\kappa)})$, i.e.

$$\varphi(\boldsymbol{\alpha}^{(\kappa+1)}, \boldsymbol{\beta}^{(\kappa+1)}) < \varphi(\boldsymbol{\alpha}^{(\kappa)}, \boldsymbol{\beta}^{(\kappa)}) \quad (25)$$

whenever $(\boldsymbol{\alpha}^{(\kappa+1)}, \boldsymbol{\beta}^{(\kappa+1)}) \neq (\boldsymbol{\alpha}^{(\kappa)}, \boldsymbol{\beta}^{(\kappa)})$.

Proof. Note that the convex function $\varphi^{(\kappa)}$ agrees with the nonconvex function φ at $(\boldsymbol{\alpha}^{(\kappa)}, \boldsymbol{\beta}^{(\kappa)})$, which is also feasible for (4), (6). Therefore

$$\begin{aligned} \varphi(\boldsymbol{\alpha}^{(\kappa+1)}, \boldsymbol{\beta}^{(\kappa+1)}) &\leq \varphi^{(\kappa)}(\boldsymbol{\alpha}^{(\kappa+1)}, \boldsymbol{\beta}^{(\kappa+1)}) \\ &< \varphi^{(\kappa)}(\boldsymbol{\alpha}^{(\kappa)}, \boldsymbol{\beta}^{(\kappa)}) \\ &= \varphi(\boldsymbol{\alpha}^{(\kappa)}, \boldsymbol{\beta}^{(\kappa)}), \end{aligned}$$

showing (25). \square

We now show that the convex program (24) admits the optimal solution in closed-form. Indeed, (24) boils down to

$$\min_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \sum_{j=1}^M \left(\frac{a_j^{(\kappa)}}{\alpha_j} + \frac{b_j^{(\kappa)}}{\beta_j} + \frac{c_j^{(\kappa)}}{2\alpha_j^2} + \frac{d_j^{(\kappa)}}{2\beta_j^2} \right) \quad \text{s.t.} \quad (4), (6) \quad (26)$$

with

$$\begin{aligned} a_j^{(\kappa)} &= \rho_j^{(\kappa)} r_j / p_j, b_j^{(\kappa)} = \rho_j^{(\kappa)} q_j / p_j, \\ c_j^{(\kappa)} &= \rho_j^{(\kappa)} \sigma_j \alpha_j^{(\kappa)} / (p_j \beta_j^{(\kappa)}), d_j^{(\kappa)} = \rho_j^{(\kappa)} \sigma_j \beta_j^{(\kappa)} / (p_j \alpha_j^{(\kappa)}) \end{aligned} \quad (27)$$

By using the Lagrangian multiplier method, it can be shown that the optimal α_j and β_j are the unique positive roots of the following compressed cubic equations

$$\alpha_j^{(\kappa)} \alpha_j + c_j^{(\kappa)} = \lambda_T \mathbb{E}[|y_j|^2] \alpha_j^3, j = 1, 2, \dots, M, \quad (28)$$

$$b_j^{(\kappa)} \beta_j + d_j^{(\kappa)} = \lambda_R \beta_j^3, j = 1, 2, \dots, M, \quad (29)$$

where $\lambda_T > 0$ and $\lambda_R > 0$ such that α_j and β_j satisfy the power constraints (4) and (6) at equality sign. Accordingly,³

$$\begin{aligned} \alpha_j^{(\kappa+1)} &= \left\{ \frac{c_j^{(\kappa)}}{2\lambda_T \mathbb{E}[|y_j|^2]} + \left[\left(\frac{c_j^{(\kappa)}}{2\lambda_T \mathbb{E}[|y_j|^2]} \right)^2 \right. \right. \\ &\quad \left. \left. + \left(\frac{a_j^{(\kappa)}}{3\lambda_T \mathbb{E}[|y_j|^2]} \right)^2 \right]^{1/2} \right\}^{1/3} + \left\{ \frac{c_j^{(\kappa)}}{2\lambda_T \mathbb{E}[|y_j|^2]} \right. \\ &\quad \left. - \left[\left(\frac{c_j^{(\kappa)}}{2\lambda_T \mathbb{E}[|y_j|^2]} \right)^2 + \left(\frac{a_j^{(\kappa)}}{3\lambda_T \mathbb{E}[|y_j|^2]} \right)^2 \right]^{1/2} \right\}^{1/3}, \end{aligned} \quad (30)$$

$$\begin{aligned} \beta_j^{(\kappa+1)} &= \left\{ \frac{d_j^{(\kappa)}}{2\lambda_R} + \left[\left(\frac{d_j^{(\kappa)}}{2\lambda_R} \right)^2 + \left(\frac{b_j^{(\kappa)}}{3\lambda_R} \right)^2 \right]^{1/2} \right\}^{1/3} \\ &\quad + \left\{ \frac{d_j^{(\kappa)}}{2\lambda_R} - \left[\left(\frac{d_j^{(\kappa)}}{2\lambda_R} \right)^2 + \left(\frac{b_j^{(\kappa)}}{3\lambda_R} \right)^2 \right]^{1/2} \right\}^{1/3} \end{aligned} \quad (31)$$

³the unique positive root of cubic equation $ax^3 - cx - d = 0$ with $a > 0, c > 0, d > 0$ is $[(d/2a) + \sqrt{(d/2a)^2 + (c/3a)^2}]^{1/3} + [(d/2a) - \sqrt{(d/2a)^2 + (c/3a)^2}]^{1/3}$

Algorithm 1 Fast iterative procedure for two-hop (2H) power allocation

- 1: Initialize $\kappa := 0$ and $(\boldsymbol{\alpha}^{(0)}, \boldsymbol{\beta}^{(0)})$ feasible to (4) and (6).
- 2: **repeat** Generate a feasible solution $(\boldsymbol{\alpha}^{(\kappa+1)}, \boldsymbol{\beta}^{(\kappa+1)})$ according to formula (30) and (31).

3: **until**

$$\frac{\varphi(\boldsymbol{\alpha}^{(\kappa)}, \boldsymbol{\beta}^{(\kappa)}) - \varphi(\boldsymbol{\alpha}^{(\kappa+1)}, \boldsymbol{\beta}^{(\kappa+1)})}{\varphi(\boldsymbol{\alpha}^{(\kappa)}, \boldsymbol{\beta}^{(\kappa)})} \leq \epsilon \quad (32)$$

for a given tolerance ϵ .

- 4: Extract $(\alpha^*, \beta^*) = (\boldsymbol{\alpha}^{(\kappa)}, \boldsymbol{\beta}^{(\kappa)})$ as a suboptimal solution of the computationally intractable problem (15).

where $\lambda_T > 0$ and λ_R are chosen so that such α_j and β_j satisfy the power constraints (4) and (6) at equality sign, which can be located by the golden search. Algorithm 1 is a pseudocode for solving the nonconvex optimization problem (19), which yields a suboptimal solution of the computationally intractable problem (15). The following result is a consequence of Proposition 1 and [32].

Proposition 2: Algorithm 1 generates a sequence $\{(\boldsymbol{\alpha}^{(\kappa)}, \boldsymbol{\beta}^{(\kappa)})\}$ of improved points, which either stops at a KKT point, or its limit is a KKT point of the nonconvex problem (19).

Proof. We have shown in (25) that $\{(\boldsymbol{\alpha}^{(\kappa)}, \boldsymbol{\beta}^{(\kappa)})\}$ is a sequence of improved points for (19). Due to the constraints (4) and (6), the convergence of $\{(\boldsymbol{\alpha}^{(\kappa)}, \boldsymbol{\beta}^{(\kappa)})\}$ can be easily shown by using Cauchy's theorem. According to [32], any limit point of $\{(\boldsymbol{\alpha}^{(\kappa)}, \boldsymbol{\beta}^{(\kappa)})\}$ satisfies the KKT conditions for optimality of (19). \square

III. APPLICATIONS TO STATIC TARGET LOCALIZATION

Let us emphasise that the receiver equations (3) and (7) for the relay and FC, and the power budget constraints (4) and (6), are common for sensor networks. In this section we show how linear and nonlinear sensor input-output equations facilitate the joint GM distribution (1) and thus utilize Algorithm 1 for sensor and relay power allocation in locating a GM target.

A. Linear sensor networks

For M linear sensors observing a GM target $\mathbf{X} \sim \sum_{i=1}^L \lambda_i \mathcal{N}(\cdot; \mathbf{m}_X^{(i)}, \mathbf{C}_X^{(i)})$, the input-output equation is [22]

$$\mathbf{Y} = \mathbf{G}\mathbf{X} + \mathbf{N}_s, \quad (33)$$

where the noise $\mathbf{N}_s \sim \mathcal{N}(\cdot; 0, \mathbf{R}_n)$ is independent from \mathbf{X} . Here $\mathbf{G} \in R^{M \times N}$, so the observation \mathbf{Y} is the noisy linear combination of the target \mathbf{X} . Then it is obvious that (\mathbf{X}, \mathbf{Y}) follows the joint distribution (1) with

$$\begin{aligned} \mathbf{m}_{\mathbf{X}, \mathbf{Y}}^{(i)} &= \begin{pmatrix} \mathbf{m}_X^{(i)} \\ \mathbf{m}_Y^{(i)} \end{pmatrix} = \begin{pmatrix} \mathbf{m}_X^{(i)} \\ \mathbf{G}\mathbf{m}_X^{(i)} \end{pmatrix}, \\ \mathbf{C}^{(i)} &= \begin{pmatrix} \mathbf{C}_X^{(i)} & \mathbf{C}_{XY}^{(i)} \\ \mathbf{C}_{YX}^{(i)} & \mathbf{C}_Y^{(i)} \end{pmatrix}, \\ \mathbf{C}_{XY}^{(i)} &= (\mathbf{C}_X^{(i)} + (\mathbf{m}_X^{(i)})^2 - (\mathbf{m}_X)^2)\mathbf{G}^T, \\ \mathbf{C}_{YX}^{(i)} &= (\mathbf{C}_{XY}^{(i)})^T, \\ \mathbf{C}_Y^{(i)} &= \mathbf{G}(\mathbf{C}_X^{(i)} + (\mathbf{m}_X^{(i)})^2 - (\mathbf{m}_X)^2)\mathbf{G}^T + \mathbf{R}_n, \\ \mathbf{C}_Y &= \mathbf{G}\mathbf{C}_X\mathbf{G}^T + \mathbf{R}_n, \mathbf{C}_{YX} = \mathbf{G}\mathbf{C}_X, \\ \mathbf{C}_{XY} &= \mathbf{C}_X\mathbf{G}^T. \end{aligned} \quad (34)$$

We first consider a static target \mathbf{X} in a two-dimensional field where the target is positioned at location $(\mathbf{X}_1, \mathbf{X}_2)$. Specifically \mathbf{X} has the following prior probability distribution

$$\begin{aligned} \mathbf{X} \sim & \frac{1}{3}\mathcal{N}(\cdot; (0, 0)^T, \mathbf{I}_2) + \frac{1}{3}\mathcal{N}(\cdot; (5, 5)^T, \mathbf{I}_2) \\ & + \frac{1}{3}\mathcal{N}(\cdot; -(5, 5)^T, \mathbf{I}_2), \end{aligned} \quad (35)$$

i.e. the target can be either located at $(0m, 0m)$, $(5m, 5m)m$ or $-(5m, 5m)$ and uncertainty in its position is characterized by a variance of $1m$.⁴ The sensor measuring parameter \mathbf{G} in (33) is determined by linearizing the following nonlinear ranging and bearing function at the target mean \mathbf{m}_X

$$g_j(\mathbf{X}) = \begin{pmatrix} \sqrt{(s_{j,x} - \mathbf{X}_1)^2 + (s_{j,y} - \mathbf{X}_2)^2} \\ \frac{s_{j,y} - \mathbf{X}_2}{s_{j,x} - \mathbf{X}_1} \end{pmatrix}, \quad (36)$$

with sensor position $(s_{j,x}, s_{j,y})^T$, $j = 1, \dots, M$.

We let the relay and FC be positioned at $(100m, 0m)$ and $(200m, 0m)$ but $M = 10$ sensors be positioned randomly surrounding the mean \mathbf{m}_X of the target. The channel gains h_{jR} and h_{jD} are determined according to the free-space path gain [33] $h = G_t G_r (\lambda/4\pi d)^2$, with the distance between two ends d , signal wavelength λ and antenna gains $G_t = 2\text{dB}$, $G_r = 5\text{dB}$. The covariance matrices are defined as $\mathbf{R}_n = \mathbf{R}_{wR} = \mathbf{R}_{wD} = 0.5\mathbf{I}$ and the sensor transmit power budget varies as $P_T = 0.1\ell$, $\ell = 1, 2, \dots, 10$ but the relay power budget is fixed at $P_R = 5$. The simulation is validated via $N_{mc} = 10000$ Monte Carlo channel realizations. To show the viability of our proposed suboptimal solution by Algorithm 1, in Fig. 2 the normalized mean squared error (NMSE)⁵ is benchmarked with three other power allocations: (i) the proposed suboptimal power allocation (2H-Algorithm 1); (ii) sensor power allocation in one-hop (1H) communication between the sensors and FC (1H-Algorithm 2), which is based on Algorithm 2 given in Appendix B; and (iii) equal power allocation schemes for one-hop (1H) and two-hop (2H) communication environments (1H and 2H-equal power). Overall, 2H-Algorithm 1 provides the lowest NMSE for all power budgets P_T . The average iterations for 2H Algorithm 1 and 1H Algorithm 2 under error tolerance $\epsilon = 10^{-3}$ in the stopping condition are shown in Table I. Additionally Fig. 3 includes two plots. The first plot demonstrates the value of the surrogate function $\varphi(\boldsymbol{\alpha}^{(\kappa)}, \boldsymbol{\beta}^{(\kappa)})$ in (18), and the second plot shows the corresponding MSE (13). Of course, from (16), the improvements of the former and the latter are not necessarily parallel and that's why the optimal solution of the former is only a suboptimal solution of the latter.

B. Nonlinear sensor networks

Rather than using (33), the input-output equation of a NSN is given as

$$\mathbf{Y} = g(\mathbf{X}) + \mathbf{N}_s. \quad (37)$$

⁴Here and after, m is referred to the conventional base unit of length, which is different from the notations \mathbf{m}_X , \mathbf{m}_Y and $\mathbf{m}_{X,Y}$ used in the paper to refer the means of random variables

⁵For an estimate $\hat{\mathbf{x}}$ of \mathbf{x} , NMSE is defined by $\|\hat{\mathbf{x}} - \mathbf{x}\|^2 / \|\mathbf{x}\|^2$

TABLE I
AVERAGE ITERATIONS OF TWO ALGORITHMS FOR LSN.

Algorithm	Power Budget P_T									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1H	4.06	4.20	4.16	4.03	3.86	3.63	3.38	3.19	3.09	3.02
2H	5.18	4.40	4.02	3.83	3.63	3.51	3.46	3.40	3.36	3.36

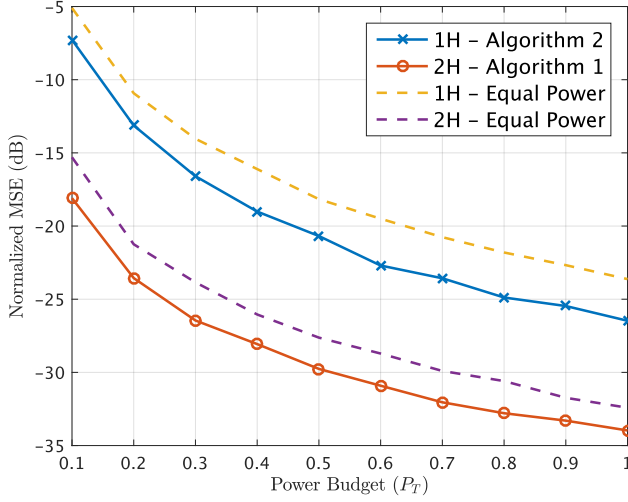


Fig. 2. Normalized MSE of LSN by different power schemes

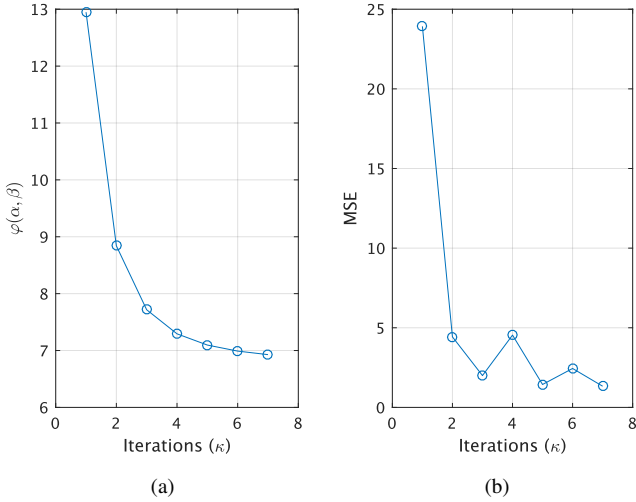


Fig. 3. (a). The function $\varphi(\alpha^{(\kappa)}, \beta^{(\kappa)})$ for power allocation at each iteration κ . (b). The MSE calculated for each estimated target $\hat{\mathbf{X}}^{(\kappa)}$ at iteration κ .

Then, (\mathbf{X}, \mathbf{Y}) approximately follows the joint GMM (1) with $\mathbf{m}_Y^{(i)}$, $\mathbf{C}_{XY}^{(i)}$ and $\mathbf{C}_Y^{(i)}$ calculated through the unscented transformation [34] as follows. For each i , take the Cholesky

decomposition⁶

$$\mathbf{C}_X^{(i)} = \sum_{r=1}^N \tilde{\mathbf{x}}^{(r)} (\tilde{\mathbf{x}}^{(r)})^T. \quad (38)$$

Accordingly, $2N + 1$ regression points $\mathbf{x}^{(r)}$, $r = 0, 1, \dots, 2N$ are defined by

$$\begin{aligned} \mathbf{x}^{(0)} &= \mathbf{m}_X^{(i)}, \\ \mathbf{x}^{(r)} &= \mathbf{m}_X^{(i)} + \sqrt{\frac{2N+1}{2}} \tilde{\mathbf{x}}^{(r)}, \\ \mathbf{x}^{(N+r)} &= \mathbf{m}_X^{(i)} - \sqrt{\frac{2N+1}{2}} \tilde{\mathbf{x}}^{(r)}, \\ &r = 1, 2, \dots, N. \end{aligned} \quad (39)$$

Clearly,

$$\begin{aligned} \mathbf{m}_X^{(i)} &= \frac{1}{2N+1} \sum_{r=0}^{2N} \mathbf{x}^{(r)}, \\ \mathbf{C}_X^{(i)} &= \frac{1}{2N+1} \sum_{r=0}^{2N} (\mathbf{x}^{(r)} - \mathbf{m}_X^{(i)})^2, \end{aligned}$$

and thereby transform $\mathbf{y}^{(r)} := g(\mathbf{x}^{(r)})$, $r = 0, 1, \dots, 2N$ for approximations

$$\begin{aligned} \mathbf{m}_Y^{(i)} &= \frac{1}{2N+1} \sum_{r=0}^{2N} \mathbf{y}^{(r)}, \\ \mathbf{C}_Y^{(i)} &= \frac{1}{2N+1} \sum_{r=0}^{2N} (\mathbf{y}^{(r)} - \mathbf{m}_Y^{(i)})^2, \\ \mathbf{C}_{XY}^{(i)} &= \frac{1}{2N+1} \sum_{r=0}^{2N} (\mathbf{x}^{(r)} - \mathbf{m}_X^{(i)}) (\mathbf{y}^{(r)} - \mathbf{m}_Y^{(i)})^T. \end{aligned} \quad (40)$$

We use nonlinear ranging and bearing functions (36) with sensor position $(s_{j,x}, s_{j,y})^T$ for $g_j(\mathbf{x}) = (g_1(\mathbf{x}), g_2(\mathbf{x}))^T$ in (37), while the target is prior characterized by (35). The simulation environment is the same as that in the previous LSN simulation. Table II shows the average iterations for 2H Algorithm 1 and 1H Algorithm 2. Fig. 4 shows the NMSE curves. Like Fig. 2, 2H provides a notable difference in dB, especially in lower fixed power budgets.

IV. DYNAMIC TARGET TRACKING BY WSN

In this section, we consider the tracking of a dynamic target, which is moving in a surveillance region. The sensor nodes are distributed geographically to take independent measurements

⁶For the SVD (singular value decomposition): $\mathbf{C}_X^{(i)} = \sum_{r=1}^N \lambda_r (\mathbf{x}^{(r)})^2$, it is obvious that $\tilde{\mathbf{x}}^{(r)} = \sqrt{\lambda_r} \mathbf{x}^{(r)}$. For notational simplicity we omit the index i in \mathbf{x}^r , i.e. rigorously speaking, it should be $\mathbf{x}^{(i,r)}$.

TABLE II
AVERAGE ITERATIONS OF TWO ALGORITHMS IN NSN.

Algorithm	Power Budget P_T									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1H	4.49	5.56	6.16	6.42	6.71	6.86	7.11	7.17	7.21	7.22
2H	5.83	6.56	6.77	6.86	6.89	6.88	6.90	6.88	6.87	6.77

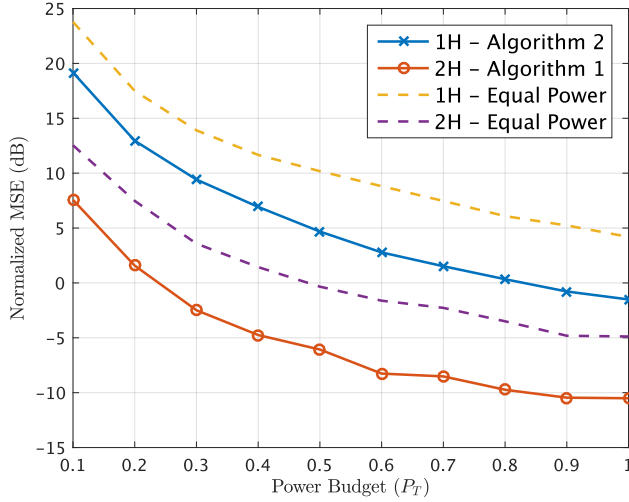


Fig. 4. Normalized MSE performance of NSN by different power schemes.

of a target's position and send these measurements to the FC via the relay node. These relayed observations are processed at each time step to update and predict the target state. The following set of equations model the considered scenario:

$$\mathbf{X}_{k+1} = \mathbf{F}_k \mathbf{X}_k + \mathbf{V}_k, \quad (41)$$

$$\mathbf{Y}_k = g_k(\mathbf{X}_k) + \mathbf{N}_k, \quad (42)$$

$$\mathbf{Z}_{k,R} = \text{diag} \left[\sqrt{\alpha_k} \sqrt{h_{jR}} \right]_{j=1,2,\dots,M} \mathbf{Y}_k + \mathbf{W}_{k,R}, \quad (43)$$

$$\mathbf{Z}_k = \mathbf{H}_{k,(\alpha,\beta)} \mathbf{Y}_k + \mathbf{W}_{k,(\alpha,\beta)}. \quad (44)$$

At time step $k = 0, 1, \dots$, (41) represents a linear Gaussian dynamical model of a target with the state transition matrix \mathbf{F}_k and (42) represents the sensor measurement, where $\mathbf{V}_k \sim \mathcal{N}(\cdot; 0, \mathbf{R}_v)$ is process noise and $\mathbf{N}_k \sim \mathcal{N}(\cdot; 0, \mathbf{R}_n)$ with diagonal \mathbf{R}_n , is the sensor measurement noise. Similarly to (3), equation (43) represents the signal received by the relay at time k , where $\sqrt{h_{jR}}$ is the channel gain between sensor j and the relay, $\mathbf{W}_{k,R} \sim \mathcal{N}(\cdot; 0, \mathbf{R}_R)$ with $\mathbf{R}_R = \text{diag}[\sigma_{jR}]_1^M$ is a corrupt noise at relay, and $\sqrt{\alpha_j}$ controls the transmitter power $P_j = \alpha_j \|y_{k,j}\|^2 = (\mathbf{C}_{\mathbf{Y}_k}(j, j) + \mathbf{m}_{\mathbf{Y}_k}^2(j))\alpha_j$ of sensor j to satisfy the fixed sum power budget $P_T > 0$, which is defined similarly to (4) as

$$\sum_{j=1}^M P_j = \sum_{j=1}^M \|y_{k,j}\|^2 \alpha_j \leq P_T. \quad (45)$$

Similar to (7), equation (44) with

$$\mathbf{H}_{k,(\alpha,\beta)} = \text{diag} \left[\sqrt{h_{jD} h_{jR} \beta_j \alpha_j / (h_{jR} \|y_{k,j}\|^2 \alpha_j + \sigma_{jR})} \right]_1^M$$

represents the signal received at the FC at time k , where $\sqrt{h_{jD}}$ is the channel gain between the relay and the FC. Note that

$$\mathbf{W}_{k,(\alpha,\beta)} = \text{diag} \left[\sqrt{h_{jD} \beta_j / (h_{jR} \|y_{k,j}\|^2 \alpha_j + \sigma_{jR})} \right]_1^M \mathbf{W}_R + \mathbf{W}_D$$

is the total noise, where $\mathbf{W}_D \sim \mathcal{N}(\cdot; 0, \text{diag}[\sigma_{jD}]_1^M)$ is the background noise, which is independent with \mathbf{Z}_R . Accordingly, $\mathbf{W}_{k,(\alpha,\beta)} \sim \mathcal{N}(\cdot; 0, \mathbf{C}_{k,(\alpha,\beta)})$ with

$$\mathbf{C}_{k,(\alpha,\beta)} = \text{diag} \left[\sigma_{jR} h_{jD} \beta_j / (h_{jR} \|y_{k,j}\|^2 \alpha_j + \sigma_{jR}) + \sigma_{jD} \right]_1^M.$$

The power levels β_j are constrained by the relay power budget P_R in (6).

Given the initial information $\mathbf{X}_{0|-1} \sim p_{\mathbf{X}_{0|-1}}(\mathbf{x}) = \sum_{i=1}^L \lambda_i(-1) \mathcal{N}(\mathbf{x}, \mathbf{m}_{\mathbf{X}_{0|-1}}^{(i)}, \mathbf{C}_{\mathbf{X}_{0|-1}}^{(i)})$, the FC iterates at time $k = 0, 1, \dots$, as followings.

- *Power constrained filtering.* Execute the unscented transformation in Subsection III.B with input $\{\lambda_i(k-1), \mathbf{m}_{\mathbf{X}_{k|k-1}}^{(i)}, \mathbf{C}_{\mathbf{X}_{k|k-1}}^{(i)}\}_{i=1}^L$, \mathbf{R}_n and $g = g_k$ to write the approximate joint GM distribution

$$(\mathbf{X}_{k|k-1}, \mathbf{Y}_k) \sim \sum_{i=1}^L \lambda_i(k-1) \mathcal{N}(\cdot, \cdot; (\mathbf{m}_{\mathbf{X}_{k|k-1}}^{(i)}, \mathbf{m}_{\mathbf{Y}_k}^{(i)}), \mathbf{C}_{k,R}^{(i)}), \quad (46)$$

where $\mathbf{C}_{k,R}^{(i)}$ is in sub-block form

$$\mathbf{C}_{k,R}^{(i)} = \begin{pmatrix} \mathbf{C}_{\mathbf{X}_{k|k-1}}^{(i)} & \mathbf{C}_{\mathbf{X}_{k|k-1} \mathbf{Y}_k}^{(i)} \\ \mathbf{C}_{\mathbf{Y}_k \mathbf{X}_{k|k-1}}^{(i)} & \mathbf{C}_{\mathbf{Y}_k}^{(i)} \end{pmatrix}.$$

Execute Algorithm 1 to output the suboptimal power allocation (α_k^*, β_k^*) and then update $\mathbf{X}_{k|k} = \mathbf{X}_{k|k-1} | \mathbf{Z}_k = \mathbf{z}_k$ as

$$\begin{aligned} \lambda_i(k) &= \lambda_i(k-1) \mathcal{N}(\mathbf{z}_k; \mathbf{H}_{\alpha_k^*, \beta_k^*} \mathbf{m}_{\mathbf{Y}_k}^{(i)}, \\ &\quad \mathbf{H}_{\alpha_k^*, \beta_k^*} \mathbf{C}_{\mathbf{Y}_k}^{(i)} \mathbf{H}_{\alpha_k^*, \beta_k^*} + \mathbf{C}_{\alpha_k^*, \beta_k^*}) \\ &\quad / \sum_{i=1}^L \lambda_i(k-1) \mathcal{N}(\mathbf{z}_k; \mathbf{H}_{\alpha_k^*, \beta_k^*} \mathbf{m}_{\mathbf{Y}_k}^{(i)}, \\ &\quad \mathbf{H}_{\alpha_k^*, \beta_k^*} \mathbf{C}_{\mathbf{Y}_k}^{(i)} \mathbf{H}_{\alpha_k^*, \beta_k^*} + \mathbf{C}_{\alpha_k^*, \beta_k^*}), \end{aligned} \quad (47)$$

$$\begin{aligned} \mathbf{m}_{\mathbf{X}_{k|k}}^{(i)} &= \mathbf{m}_{\mathbf{X}_{k|k-1}}^{(i)} + \mathbf{C}_{\mathbf{Y}_k}^{(i)T} \mathbf{H}_{\alpha_k^*, \beta_k^*} \\ &\quad \times (\mathbf{H}_{\alpha_k^*, \beta_k^*} \mathbf{C}_{\mathbf{Y}_k}^{(i)} \mathbf{H}_{\alpha_k^*, \beta_k^*} + \mathbf{C}_{\alpha_k^*, \beta_k^*})^{-1} \\ &\quad \times (\mathbf{z}_k - \mathbf{H}_{\alpha_k^*, \beta_k^*} \mathbf{m}_{\mathbf{Y}_k}^{(i)}), \end{aligned} \quad (48)$$

$$\begin{aligned} \mathbf{C}_{\mathbf{X}_{k|k}}^{(i)} &= \mathbf{C}_{\mathbf{X}_{k|k-1}}^{(i)} - \mathbf{C}_{\mathbf{X}_{k|k-1} \mathbf{Y}_k}^{(i)} \mathbf{H}_{\alpha_k^*, \beta_k^*} \\ &\quad \times (\mathbf{H}_{\alpha_k^*, \beta_k^*} \mathbf{C}_{\mathbf{Y}_k}^{(i)} \mathbf{H}_{\alpha_k^*, \beta_k^*} + \mathbf{C}_{\alpha_k^*, \beta_k^*})^{-1} \\ &\quad \times \mathbf{H}_{\alpha_k^*, \beta_k^*} (\mathbf{C}_{\mathbf{X}_{k|k-1} \mathbf{Y}_k}^{(i)})^T, \end{aligned} \quad (49)$$

$$i = 1, 2, \dots, L,$$

$$\mathbf{X}_{k|k} \sim \sum_{i=1}^L \lambda_i(k) \mathcal{N}(\cdot; \mathbf{m}_{\mathbf{X}_{k|k}}^{(i)}, \mathbf{C}_{\mathbf{X}_{k|k}}^{(i)}), \quad (50)$$

$$\mathbf{x}_{k|k} = \sum_{i=1}^L \lambda_i(k) \mathbf{m}_{\mathbf{X}_{k|k}}^{(i)}. \quad (51)$$

- *State distribution prediction.* Write the joint GM distribution

$$(\mathbf{X}_{k|k}, \mathbf{X}_{k+1|k}) \sim \sum_{i=1}^L \lambda_i(k) \mathcal{N}\left(\left(\cdot, \cdot\right); \begin{pmatrix} \mathbf{m}_{\mathbf{X}_{k|k}}^{(i)} \\ \mathbf{m}_{\mathbf{X}_{k+1|k}}^{(i)} \end{pmatrix}, \begin{pmatrix} \mathbf{C}_{\mathbf{X}_{k|k}}^{(i)} & \mathbf{C}_{\mathbf{X}_{k|k} \mathbf{X}_{k+1|k}}^{(i)} \\ \mathbf{C}_{\mathbf{X}_{k+1|k} \mathbf{X}_{k|k}}^{(i)} & \mathbf{C}_{\mathbf{X}_{k+1|k}}^{(i)} \end{pmatrix}\right), \quad (52)$$

with

$$\begin{aligned} \mathbf{m}_{\mathbf{X}_{k+1|k}}^{(i)} &= \mathbf{F}_k \mathbf{m}_{\mathbf{X}_{k|k}}^{(i)}, \\ \mathbf{C}_{\mathbf{X}_{k+1|k} \mathbf{X}_{k|k}}^{(i)} &= (\mathbf{C}_{\mathbf{X}_{k|k} \mathbf{X}_{k+1|k}}^{(i)})^T, \\ \mathbf{C}_{\mathbf{X}_{k|k} \mathbf{X}_{k+1|k}}^{(i)} &= (\mathbf{C}_{\mathbf{X}_{k|k}}^{(i)} + (\mathbf{m}_{\mathbf{X}_{k|k}}^{(i)})^2 - (\mathbf{m}_{\mathbf{X}_{k+1|k}}^{(i)})^2) \mathbf{F}_k^T, \\ \mathbf{C}_{\mathbf{X}_{k+1|k}}^{(i)} &= \mathbf{F}_k (\mathbf{C}_{\mathbf{X}_{k|k}}^{(i)} + (\mathbf{m}_{\mathbf{X}_{k|k}}^{(i)})^2 - (\mathbf{m}_{\mathbf{X}_{k+1|k}}^{(i)})^2) \mathbf{F}_k^T + \mathbf{R}_v \end{aligned}$$

to update

$$\mathbf{X}_{k+1|k} \sim \sum_{i=1}^L \lambda_i(k) \mathcal{N}(\cdot; \mathbf{m}_{\mathbf{X}_{k+1|k}}^{(i)}, \mathbf{C}_{\mathbf{X}_{k+1|k}}^{(i)}). \quad (53)$$

Thus, the track of \mathbf{X}_k is $\mathbf{x}_{k|k}$ defined by equation (51), while the track of \mathbf{X}_k by LMMSE estimate is defined by (68) (in Appendix B) with $\mathbf{X} \rightarrow \mathbf{X}_{k|k}$ and $\mathbf{Y} \rightarrow \mathbf{Z}_k$.

A. Linear Sensor Networks

Consider a scenario with a 2D dynamic target moving in a region $[-45m, -25m] \times [-50m, 100m]$, where the relay is at $(100m, 0m)$, FC is at $(200m, 0m)$ and $M = 10$ sensors are randomly distributed within a surveillance region $[-100m, 100m] \times [-100m, 100m]$. The state $\mathbf{X}_k = (p_{xk}, p_{yk}, \dot{p}_{xk}, \dot{p}_{yk})^T$ of the target consists of position (p_{xk}, p_{yk}) and velocity $(\dot{p}_{xk}, \dot{p}_{yk})$, while the measurement is a noise corrupted version of the position. The target's dynamics follow the linear Gaussian dynamical model (41) with

$$\mathbf{F}_k = \begin{pmatrix} \mathbf{I}_2 & T\mathbf{I}_2 \\ 0_2 & \mathbf{I}_2 \end{pmatrix}, \mathbf{R}_v = \begin{pmatrix} \frac{T^4}{4}\mathbf{I}_2 & \frac{T^3}{2}\mathbf{I}_2 \\ \frac{T^3}{2}\mathbf{I}_2 & T^2\mathbf{I}_2 \end{pmatrix},$$

where $T = 1s$ is the sampling period. All the sensors are linear, so (42) is

$$g_k(\mathbf{X}_k) = \begin{pmatrix} \mathbf{I}_2 & 0_{2 \times 2} \end{pmatrix} \mathbf{X}_k.$$

We assume that the initial state of the target is

$$\mathbf{X}_{0|-1} \sim \frac{1}{3} \sum_{i=1}^3 \mathcal{N}(x; \mathbf{m}_{\mathbf{X}_{0|-1}}^{(i)}, \mathbf{C}_{\mathbf{X}_{0|-1}}^{(i)}),$$

where $\mathbf{m}_{\mathbf{X}_{0|-1}}^{(1)} = (0, 0, 0, 0)^T$, $\mathbf{m}_{\mathbf{X}_{0|-1}}^{(2)} = (50, 50, 0, 0)^T$, $\mathbf{m}_{\mathbf{X}_{0|-1}}^{(3)} = (-50, -50, 0, 0)^T$, $\mathbf{C}_{\mathbf{X}_{0|-1}}^{(1)} = \text{diag}[1, 1, 2, 2]$, $\mathbf{C}_{\mathbf{X}_{0|-1}}^{(2)} = \text{diag}[1, 2, 3, 4]$, and $\mathbf{C}_{\mathbf{X}_{0|-1}}^{(3)} = \text{diag}[2, 3, 3, 4]$.

Fig. 5 shows the motion of the target, with the initial position and final position at approximately $[-40m, -40m]$ and $[-30m, 83m]$, respectively. Overall, all the Algorithms 1 and 2 help track the true path considerably well because the blue line (true track) is hidden by the estimated tracks. This implies the Algorithms can accurately recognize the true track, however at larger distances from the Relay (or FC), 2H Algorithm 1 is more robust in tracking the path of the target, followed by 2H using equal power allocation (2H-EP), 1H and 1H with equal power allocation (1H-EP).

The MSE for the estimated positions is shown in Fig. 6. In the majority of time steps, the 2H Algorithm 1 has less MSE in comparison to the other methods. Furthermore, it can also be seen that 2H-EP even outperforms both 1H and 1H-EP in most time steps. Thus it can be concluded that 2H completely outperforms 1H for this example.

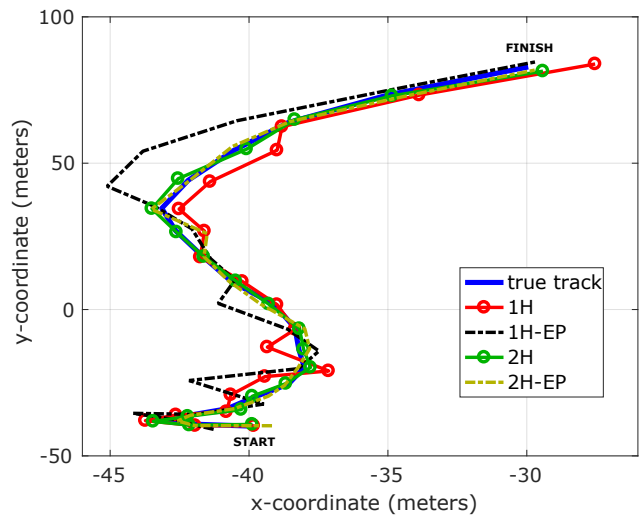


Fig. 5. Path of a constant velocity target and the estimated tracks.

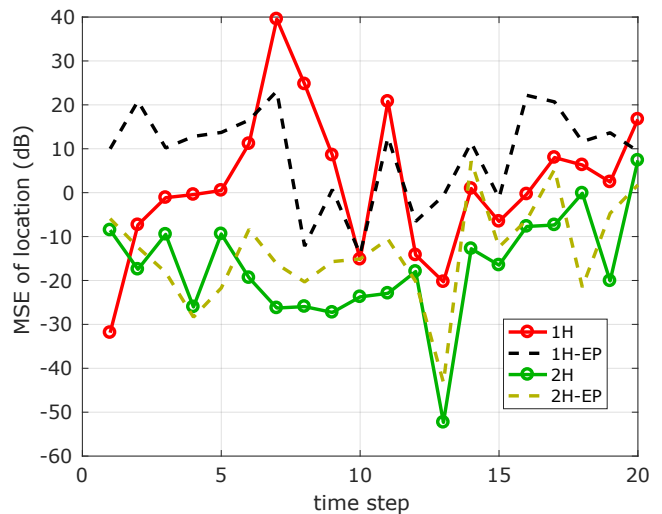


Fig. 6. Comparison of MSE performance for a dynamic target in a LSN.

B. Nonlinear Sensor Networks

In this example, we study the tracking performance of a manoeuvring target in a region $[0m, 150m] \times [0m, 150m]$. The sensors are distributed geographically over a surveillance region $[-150m, 150m] \times [-150m, 150m]$ to take maximum advantage of estimation diversity. The relay and FC are positioned at $(200m, 0m)$ and $(400m, 0m)$, respectively. A coordinated turn model (see e.g. [35]) characterizes the dynamics of the target. The target kinematic state $\mathbf{X}_k = (p_{x_k}, \dot{p}_{x_k}, p_{y_k}, \dot{p}_{y_k})^T$ consists of the target position (p_x, p_y) and its velocity (\dot{p}_x, \dot{p}_y) . The state dynamical model of the target is assumed to be linear Gaussian, which is mathematically expressed by (41) with

$$\mathbf{F}_k = \begin{pmatrix} 1 & \frac{\sin \omega T}{\omega} & 0 & -\frac{1 - \cos \omega T}{\omega} \\ 0 & \cos \omega T & 0 & -\sin \omega T \\ 0 & \frac{1 - \cos \omega T}{\omega} & 1 & \frac{\sin \omega T}{\omega} \\ 0 & \sin \omega T & 0 & \cos \omega T \end{pmatrix},$$

$$\mathbf{R}_v = \begin{pmatrix} \frac{T^4}{4} & \frac{T^3}{2} & 0 & 0 \\ \frac{T^3}{2} & T^2 & 0 & 0 \\ 0 & 0 & \frac{T^4}{4} & \frac{T^3}{2} \\ 0 & 0 & \frac{T^3}{2} & T^2 \end{pmatrix},$$

where T is the sampling period and ω is the turn rate of the manoeuvring target. The sensor nonlinear measurements include range and bearing information of the vehicle, which is represented by (42) with

$$g_k(\mathbf{X}) = \left(\left(\begin{array}{c} \sqrt{(s_{1,x} - \mathbf{X}_k(1))^2 + (s_{1,y} - \mathbf{X}_k(3))^2} \\ \frac{s_{1,y} - \mathbf{X}_k(3)}{s_{1,x} - \mathbf{X}_k(1)} \end{array} \right), \dots, \left(\begin{array}{c} \sqrt{(s_{M,x} - \mathbf{X}_k(1))^2 + (s_{M,y} - \mathbf{X}_k(3))^2} \\ \frac{s_{M,y} - \mathbf{X}_k(3)}{s_{M,x} - \mathbf{X}_k(1)} \end{array} \right) \right),$$

where $(s_{j,x}, s_{j,y})$ is the position of sensor j . The initial state of target is assumed as $\mathbf{X}_{0|0} \sim 0.5 \sum_{i=1}^2 \mathcal{N}(x; \mathbf{m}_{\mathbf{X}_{0|0}}^{(i)}, \mathbf{C}_{\mathbf{X}_{0|0}}^{(i)})$, where $\mathbf{m}_{\mathbf{X}_{0|0}}^{(1)} = (10, 5, 10, 5)^T$, $\mathbf{m}_{\mathbf{X}_{0|0}}^{(2)} = (20, 5, 30, 5)^T$, $\mathbf{C}_{\mathbf{X}_{0|0}}^{(i)} \equiv \text{diag}[3, 3, 2, 2]$.

Fig. 7 shows the two-dimensional motion of a manoeuvring target. At $k = 1s$, the target begins its motions from a position $(9m, 11m)$ at a constant velocity of $5m/s$ and after $14s$ performs a counterclockwise turn for $11s$ at a turn rate of $\omega = 0.2rad/s$. It then takes a clockwise turn after $25s$ with a turn rate of $\omega = -0.1rad/s$ until it reaches its final position at $(110m, 125m)$. The 2H Algorithm estimates the true path of the target very accurately in comparison to the other Algorithms. It can be seen that 1H and 1H-EP perform poorly since at several time steps their estimated paths diverge from the original path of the target.

A comparison of MSE for target locations can be seen in Fig. 8. Both 1H and 1H-EP have a significantly higher error overall in comparison to the 2H Algorithms. The 2H Algorithm has quite lower MSE in comparison to 2H-EP, which clearly demonstrates the advantage of 2H over the other methods.

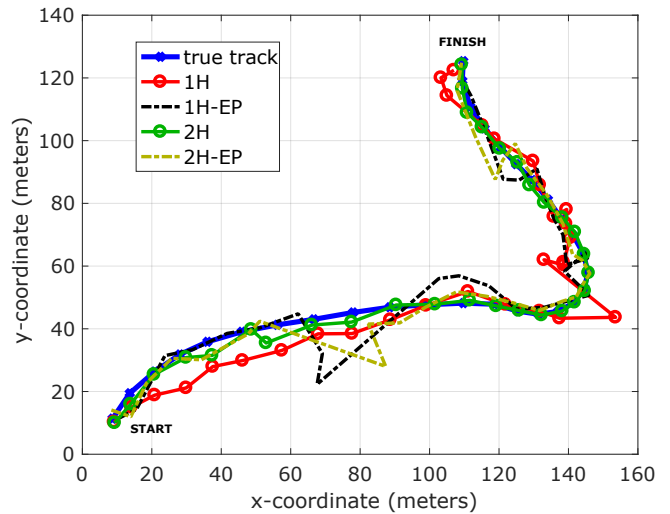


Fig. 7. Path of a manoeuvring target and the estimated tracks.

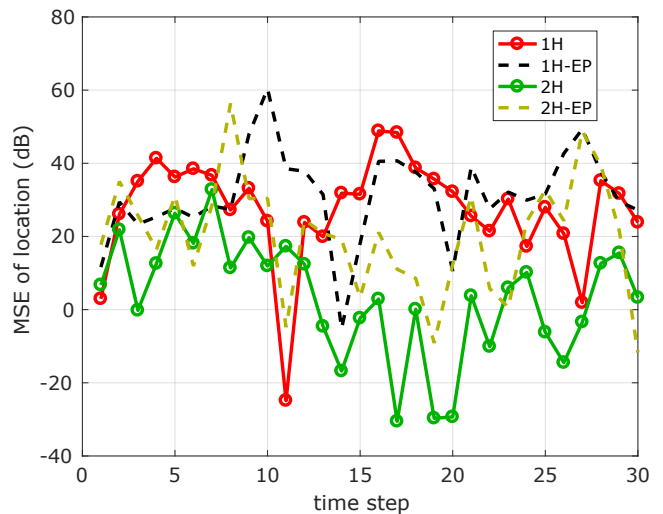


Fig. 8. Comparison of MSE performance for a dynamic target in a NSN.

C. Linear sensor networks for nonlinear dynamics

The unscented transformation in subsection III.B can be applied to a target with nonlinear dynamics to have an approximated joint GM distribution (52) for state prediction, which however may not lead to tracking a target's true path. Consider a typical third-order nonlinear autoregressive model described mathematically by [22], [28] as

$$\mathbf{q}_{k+2} = -0.1\mathbf{q}_{k+1} - \mathbf{q}_k^3 + \mathbf{v}_k$$

with the noise corrupted observations $\mathbf{y}_k^{(i)} = g^{(i)}\mathbf{q}_k + \mathbf{n}_k^{(i)}$, where $\mathbf{v}_k \sim \mathcal{N}(\cdot, 0, 0.04)$, $\mathbf{n}_k^{(i)} \sim \mathcal{N}(\cdot, 0, 0.1)$ and $g^{(i)} = 1 + 0.11(\ell - 1)$, $\ell = 1, 2, \dots, 10$.

By choosing the state $\mathbf{X}_k = (\mathbf{q}_k, \mathbf{q}_{k+1})^T \in \mathbb{R}^2$, the state dynamic and measurement equations are

$$\mathbf{X}_{k+1} = f(\mathbf{X}_k) + \mathbf{V}, \mathbf{Y}_k = \mathbf{G}\mathbf{X}_k + \mathbf{N}_s \quad (54)$$

where

$$f(\mathbf{X}_k) = \begin{pmatrix} \mathbf{X}_k(2) \\ -\mathbf{X}_k(1)^3 - 0.1\mathbf{X}_k(2) \end{pmatrix}, \mathbf{V} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \mathbf{v}_k,$$

$$\mathbf{G} = \begin{pmatrix} g^{(1)} & 0 \\ g^{(2)} & 0 \\ \dots & \dots \\ g^{(M)} & 0 \end{pmatrix}, \mathbf{n} = (\mathbf{n}^{(1)}, \dots, \mathbf{n}^{(M)})^T.$$

The initial state of target is given by

$$\mathbf{X}_{0|-1} \sim 0.5 \sum_{i=1}^2 \mathcal{N}(x; \mathbf{m}_{\mathbf{X}_{0|-1}}^{(i)}, \mathbf{C}_{\mathbf{X}_{0|-1}}^{(i)}),$$

where $\mathbf{m}_{\mathbf{X}_{0|-1}}^{(1)} = (0.1, 0.1)^T$, $\mathbf{m}_{\mathbf{X}_{0|-1}}^{(2)} = -(0.1, 0.1)^T$ and $\mathbf{C}_{\mathbf{X}_{0|-1}}^{(i)} = \begin{pmatrix} 1 & \rho^{(i)} \\ \rho^{(i)} & 1 \end{pmatrix}$, $\rho^{(1)} = 0.75$, $\rho^{(2)} = 0.8$. In this example, using the unscented transformation for updating the joint GM distribution of $\mathbf{X}_{k|k}$ and $\mathbf{X}_{k+1|k}$ for state prediction will not track the target. Following [28], we represent

$$\mathbf{X}_{k+1} = (F + B(\mathbf{I}_2 - \Delta(\mathbf{X}_k)D)^{-1}\Delta(\mathbf{X}_k)C)\mathbf{X}_k$$

for

$$F = \begin{pmatrix} 0 & 1 \\ 0 & -0.1 \end{pmatrix}, B = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix},$$

$$D = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, C = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \Delta(\mathbf{x}_k) = \mathbf{x}_k(1)\mathbf{I}_2$$

and use the following procedure for updating $\mathbf{C}_{\mathbf{X}_{k+1|k}}^{(i)}$ and $\mathbf{m}_{\mathbf{X}_{k+1|k}}^{(i)}$ from $\mathbf{C}_{\mathbf{X}_{k|k}}^{(i)}$ and $\mathbf{m}_{\mathbf{X}_{k|k}}^{(i)}$.

- Take the Cholesky decomposition $\mathbf{C}_{\mathbf{X}_{k|k}}^{(i)} = (\tilde{x}^{(1)})^2 + (\tilde{x}^{(2)})^2$ and set $x^{(0)} = \mathbf{m}_{\mathbf{X}_{k|k}}^{(i)}$, $x^{(r)} = x^{(0)} + \sqrt{5/2}\tilde{x}^{(r)}$ and $x^{(r+2)} = x^{(0)} - \sqrt{5/2}\tilde{x}^{(r)}$, $r = 1, 2$.
- Set $\bar{w}_\Delta = \frac{1}{5}(\mathbf{I}_2 - \Delta(x^{(0)})D)^{-1} \sum_{r=0}^4 \Delta(x^{(r)})Cx^{(r)}$ and then $w_{\Delta r} = \Delta(x^{(r)})(Cx^{(r)} + D\bar{w}_\Delta)$, $r = 0, 1, \dots, 5$.
- Set $R_\Delta = \frac{1}{5} \sum_{r=0}^4 (w_{\Delta r} - \bar{w}_\Delta)(w_{\Delta r} - \bar{w}_\Delta)^T$ and take

$$\mathbf{m}_{\mathbf{X}_{k+1|k}}^{(i)} = F\mathbf{m}_{\mathbf{X}_{k|k}}^{(i)} + B\bar{w}_\Delta,$$

$$\mathbf{C}_{\mathbf{X}_{k+1|k}}^{(i)} = F\mathbf{C}_{\mathbf{X}_{k|k}}^{(i)}F^T + BR_\Delta B^T + \begin{pmatrix} 0 & 0 \\ 0 & \sigma_{\mathbf{v}_k} \end{pmatrix}.$$

The trajectory of $\mathbf{X}_k(1)$ for 50 time steps along with the estimated tracks are shown in Fig. 9 and the MSE is plotted in Fig. 10. Although it may be difficult to visually compare the trajectories in Fig. 9, the MSE plot in Fig. 10 clearly shows that the 2H Algorithm outperforms the other Algorithms for the measured state estimation. In most time steps, the optimized 2H power allocation Algorithm offers less MSE compared to 2H equal power allocation, as well as both 1H-based methods.

V. CONCLUSION

The paper addressed the problem of joint sensor and relay power allocation for locating a stationary Gaussian mixture target or for tracking a dynamic Gaussian mixture target by either linear sensor networks or nonlinear sensor networks. We considered scenarios where the sensors' noisy observations are transmitted to the relay that amplifies and then forwards them

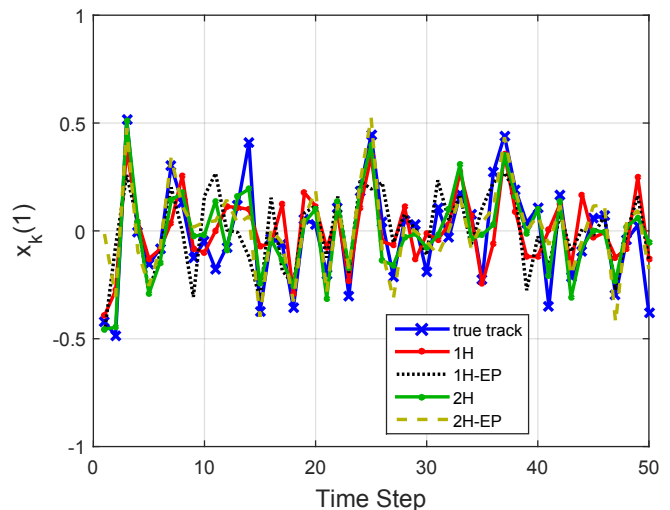


Fig. 9. The true and estimated trajectory of the state \mathbf{x}_k .

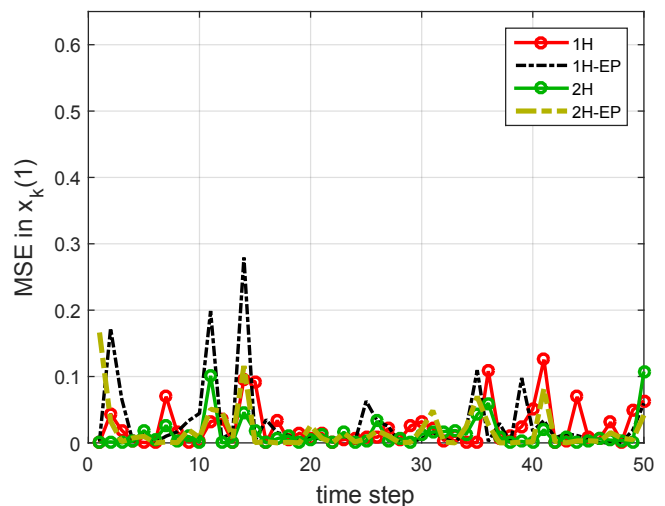


Fig. 10. Comparison of MSE at each time step.

to the FC. To arrive at an accurate estimate of a targets state, a novel technique based on tractable and scalable optimization was proposed to optimize Bayesian filtering under low sensor transmitter and relay power budgets. Numerical examples have confirmed the merits of our proposed technique.

APPENDIX A: FUNDAMENTAL MATRIX INEQUALITIES FOR GMM

A Gaussian mixture random variable is characterized by a PDF in the form,

$$p_{\mathbf{X}}(\mathbf{x}) = \sum_{i=1}^L \lambda_i \mathcal{N}(\mathbf{x}; \mathbf{m}_{\mathbf{X}}^{(i)}, \mathbf{C}_{\mathbf{X}}^{(i)}), \sum_{i=1}^L \lambda_i = 1, \lambda_i > 0. \quad (55)$$

This PDF is a weighted sum of L component Gaussian PDFs $p_{\mathbf{X}^{(i)}}(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \mathbf{m}_{\mathbf{X}}^{(i)}, \mathbf{C}_{\mathbf{X}}^{(i)})$. Here $\mathbf{m}_{\mathbf{X}}^{(i)}$ is the statistical mean and $\mathbf{C}_{\mathbf{X}}^{(i)}$ is the covariance matrix defined by $\mathbf{C}_{\mathbf{X}}^{(i)} = \mathbb{E}[(\mathbf{X} - \mathbf{m}_{\mathbf{X}}^{(i)})^2]$. By straightforward calculation, the mean

vector and auto-covariance matrix of such a random variable is

$$\mathbf{m}_X = \sum_{i=1}^L \lambda_i \mathbf{m}_X^{(i)}, \quad (56)$$

$$\mathbf{C}_X = \sum_{i=1}^L \lambda_i \left[\mathbf{C}_X^{(i)} + (\mathbf{m}_X^{(i)})^2 \right] - (\mathbf{m}_X)^2 \quad (57)$$

$$= \sum_{i=1}^L \lambda_i \mathbf{R}_X^{(i)} - (\mathbf{m}_X)^2. \quad (58)$$

The last equality implies

$$\mathbf{R}_X = \sum_{i=1}^L \lambda_i \mathbf{R}_X^{(i)}. \quad (59)$$

It is worth noticing the following convex matrix equality

$$\begin{aligned} \sum_{i=1}^L \lambda_i (\mathbf{m}_X^{(i)})^2 - (\mathbf{m}_X)^2 &= \\ \sum_{i=1}^L \lambda_i (1 - \lambda_i) (\mathbf{m}_X^{(i)})^2 & \\ - \sum_{i=1}^L \sum_{1 \leq j \leq L, j \neq i} \lambda_i \lambda_j \mathbf{m}_X^{(i)} (\mathbf{m}_X^{(j)})^T &= \\ \sum_{i=1}^L \lambda_i \left(\sum_{1 \leq j \leq L, j \neq i} \lambda_j \right) (\mathbf{m}_X^{(i)})^2 & \\ - \sum_{i=1}^L \sum_{1 \leq j \leq L, j \neq i} \lambda_i \lambda_j \mathbf{m}_X^{(i)} (\mathbf{m}_X^{(j)})^T &= \\ \sum_{i=1}^L \sum_{1 \leq j \leq L, j \neq i} \lambda_i \lambda_j [(\mathbf{m}_X^{(i)})^2 + (\mathbf{m}_X^{(j)})^2 & \\ - \mathbf{m}_X^{(i)} (\mathbf{m}_X^{(j)})^T - \mathbf{m}_X^{(j)} (\mathbf{m}_X^{(i)})^T] &= \\ \sum_{i=1}^L \sum_{1 \leq j \leq L, j \neq i} \lambda_i \lambda_j (\mathbf{m}_X^{(i)} - \mathbf{m}_X^{(j)})^2 &\succeq 0, \quad (60) \end{aligned}$$

which together with (57) gives the following bound

$$\mathbf{C}_X \succeq \sum_{i=1}^L \lambda_i \mathbf{C}_X^{(i)}. \quad (61)$$

One of the most important features of a Gaussian PDF is its factorized representation (see e.g. [36, Th. 2.1]) $\mathcal{N}(\mathbf{x}, \mathbf{y}; \mathbf{m}_{\mathbf{X}, \mathbf{Y}}, \mathbf{C}) = \mathcal{N}(\mathbf{x}; \mathbf{m}_{\mathbf{X}|\mathbf{Y}}, \mathbf{C}_{\mathbf{X}|\mathbf{Y}}) \mathcal{N}(\mathbf{y}; \mathbf{m}_{\mathbf{Y}}, \mathbf{C}_{\mathbf{Y}})$ for $\mathbf{m}_{\mathbf{X}, \mathbf{Y}} := \begin{pmatrix} \mathbf{m}_X \\ \mathbf{m}_Y \end{pmatrix}$, $\mathbf{C} = \begin{pmatrix} \mathbf{C}_X & \mathbf{C}_{\mathbf{X}\mathbf{Y}} \\ \mathbf{C}_{\mathbf{Y}\mathbf{X}} & \mathbf{C}_Y \end{pmatrix}$, $\mathbf{m}_{\mathbf{X}|\mathbf{Y}} = \mathbf{m}_X + \mathbf{C}_{\mathbf{X}\mathbf{Y}}(\mathbf{C}_Y)^{-1}(\mathbf{y} - \mathbf{m}_Y)$, $\mathbf{C}_{\mathbf{X}|\mathbf{Y}} = \mathbf{C}_X - \mathbf{C}_{\mathbf{X}\mathbf{Y}}(\mathbf{C}_Y)^{-1}\mathbf{C}_{\mathbf{Y}\mathbf{X}}$. This also means $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = \mathcal{N}(\mathbf{x}, \mathbf{m}_{\mathbf{X}|\mathbf{Y}}, \mathbf{C}_{\mathbf{X}|\mathbf{Y}})$, i.e. the conditional $\mathbf{X}|\mathbf{Y}$ of two jointly Gaussian RVs \mathbf{X} and \mathbf{Y} is still jointly Gaussian. Using this we can state the following result (see e.g. [25]):

Theorem 2: Suppose (\mathbf{X}, \mathbf{Y}) is a jointly Gaussian mixture RV characterized by (1). Then

$$f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{Y}}(\mathbf{y}) \sum_{i=1}^L \lambda_i(\mathbf{y}) \mathcal{N}(\mathbf{x}, \mathbf{m}_{\mathbf{X}^{(i)}|\mathbf{Y}^{(i)}}, \mathbf{C}_{\mathbf{X}^{(i)}|\mathbf{Y}^{(i)}}) \quad (62)$$

for

$$\begin{aligned} \mathbf{m}_{\mathbf{X}^{(i)}|\mathbf{Y}^{(i)}} &:= \mathbf{m}_X^{(i)} + \mathbf{C}_{\mathbf{X}\mathbf{Y}}(\mathbf{C}_Y^{(i)})^{-1}(\mathbf{y} - \mathbf{m}_Y^{(i)}), \\ \mathbf{C}_{\mathbf{X}^{(i)}|\mathbf{Y}^{(i)}} &:= \mathbf{C}_X^{(i)} - \mathbf{C}_{\mathbf{X}\mathbf{Y}}(\mathbf{C}_Y^{(i)})^{-1}\mathbf{C}_{\mathbf{Y}\mathbf{X}}^{(i)T}, \\ f_{\mathbf{Y}}(\mathbf{y}) &:= \sum_{i=1}^L \lambda_i \mathcal{N}(\mathbf{y}; \mathbf{m}_Y^{(i)}, \mathbf{C}_Y^{(i)}), \end{aligned}$$

and

$$\lambda_i(\mathbf{y}) := \lambda_i \mathcal{N}(\mathbf{y}; \mathbf{m}_Y^{(i)}, \mathbf{C}_Y^{(i)}) / f_{\mathbf{Y}}(\mathbf{y}).$$

□

It follows from the above Theorem that $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^L \lambda_i(\mathbf{y}) \mathcal{N}(\mathbf{x}, \mathbf{m}_{\mathbf{X}^{(i)}|\mathbf{Y}^{(i)}}, \mathbf{C}_{\mathbf{X}^{(i)}|\mathbf{Y}^{(i)}})$, which is Gaussian mixture in \mathbf{x} . The MMSE estimate for \mathbf{X} based on the measurement $\mathbf{Y} = \mathbf{y}$ is

$$\hat{\mathbf{x}}_{mmse} = \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}] := \int \mathbf{x} f_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x}) d\mathbf{x} \quad (63)$$

$$= \sum_{i=1}^L \lambda_i(\mathbf{y}) \mathbf{m}_{\mathbf{X}^{(i)}|\mathbf{Y}^{(i)}}. \quad (64)$$

The covariance matrix $\mathbf{C}_{mmse}(\mathbf{y})$ of the estimation error is equal to the conditional covariance matrix of $\mathbf{C}_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}$ of RV $\mathbf{X}|\mathbf{Y} = \mathbf{y}$:

$$\begin{aligned} \mathbf{C}_{mmse}(\mathbf{y}) &= \mathbf{C}_{\mathbf{X}|\mathbf{Y}=\mathbf{y}} \\ &:= \int (\mathbf{x} - \hat{\mathbf{x}}_{mmse})^2 f_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x}) d\mathbf{x} \quad (65) \\ &= \sum_{i=1}^L \lambda_i(\mathbf{y}) (\mathbf{C}_{\mathbf{X}^{(i)}|\mathbf{Y}^{(i)}} + (\mathbf{m}_{\mathbf{X}^{(i)}|\mathbf{Y}^{(i)}})^2) \\ &\quad - (\hat{\mathbf{x}}_{mmse})^2. \quad (66) \end{aligned}$$

MMSE of Bayesian estimate $\mathbb{E}[\|\mathbf{X} - \mathbf{X}|\mathbf{Y} = \mathbf{y}\|^2]$ for \mathbf{X} based on observation $\mathbf{Y} = \mathbf{y}$ is thus

$$\begin{aligned} \epsilon_{mmse}^2(\mathbf{y}) &= \text{Trace}(\mathbf{C}_{mmse}(\mathbf{y})) \\ &= \sum_{i=1}^L \lambda_i(\mathbf{y}) [\text{Trace}(\mathbf{C}_{\mathbf{X}^{(i)}|\mathbf{Y}^{(i)}}) \\ &\quad + \|\mathbf{m}_{\mathbf{X}^{(i)}|\mathbf{Y}^{(i)}}\|^2] - \|\hat{\mathbf{x}}_{mmse}\|^2. \quad (67) \end{aligned}$$

On the other hand, by [28, Theorem 1], the LMMSE estimate for \mathbf{X} based on observation $\mathbf{Y} = \mathbf{y}$ is

$$\hat{\mathbf{x}}_{lmse} = \mathbf{m}_X + \mathbf{C}_{\mathbf{Y}\mathbf{X}}^T \mathbf{C}_Y^{-1} (\mathbf{y} - \mathbf{m}_Y) \quad (68)$$

with MSE covariance

$$\mathbf{C}_{lmse} = \int (\mathbf{x} - \hat{\mathbf{x}}_{lmse})^2 f_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x}) d\mathbf{x} \quad (69)$$

$$= \mathbf{C}_X - \mathbf{C}_{\mathbf{X}\mathbf{Y}} \mathbf{C}_Y^{-1} \mathbf{C}_{\mathbf{Y}\mathbf{X}}, \quad (70)$$

and MSE $\epsilon_{lmse}^2 = \text{Trace}(\mathbf{C}_{lmse})$ for all \mathbf{y} . Here, according to (56), (57) and (60)

$$\mathbf{m}_{\mathbf{X},\mathbf{Y}} := \begin{pmatrix} \mathbf{m}_{\mathbf{X}} \\ \mathbf{m}_{\mathbf{Y}} \end{pmatrix} = \sum_{i=1}^L \lambda_i \mathbf{m}_{\mathbf{X},\mathbf{Y}}^{(i)} \quad \text{for } \mathbf{m}_{\mathbf{X},\mathbf{Y}}^{(i)} := \begin{pmatrix} \mathbf{m}_{\mathbf{X}}^{(i)} \\ \mathbf{m}_{\mathbf{Y}}^{(i)} \end{pmatrix}, \quad (71)$$

$$\begin{pmatrix} \mathbf{C}_{\mathbf{X}} & \mathbf{C}_{\mathbf{X}\mathbf{Y}} \\ \mathbf{C}_{\mathbf{Y}\mathbf{X}} & \mathbf{C}_{\mathbf{Y}} \end{pmatrix} = \sum_{i=1}^L \lambda_i (\mathbf{C}^{(i)} + (\mathbf{m}_{\mathbf{X},\mathbf{Y}}^{(i)})^2) - (\mathbf{m}_{\mathbf{X},\mathbf{Y}})^2 \quad (72)$$

$$\succeq \mathbf{C}(\lambda), \quad (73)$$

for

$$\mathbf{C}(\lambda) := \begin{pmatrix} \mathbf{C}_{\mathbf{X}}(\lambda) & \mathbf{C}_{\mathbf{X}\mathbf{Y}}(\lambda) \\ \mathbf{C}_{\mathbf{Y}\mathbf{X}}(\lambda) & \mathbf{C}_{\mathbf{Y}}(\lambda) \end{pmatrix} = \sum_{i=1}^L \lambda_i \mathbf{C}^{(i)}.$$

Lemma 1: (Shur's convex and monotonic inequalities) For all matrices $\mathbf{C}^{(i)} \succeq 0$ and $\sum_{i=1}^L \lambda_i = 1, \lambda_i \geq 0$ the following convex matrix inequality and monotonic matrix inequality hold true:

$$\mathbf{C}_{\mathbf{X}\mathbf{Y}}(\lambda)(\mathbf{C}_{\mathbf{Y}}(\lambda))^{-1}\mathbf{C}_{\mathbf{Y}\mathbf{X}}(\lambda) \preceq \sum_{i=1}^L \lambda_i \mathbf{C}_{\mathbf{X}\mathbf{Y}}^{(i)}(\mathbf{C}_{\mathbf{Y}}^{(i)})^{-1}\mathbf{C}_{\mathbf{Y}\mathbf{X}}^{(i)}, \quad (74)$$

and

$$\sum_{i=1}^L (\mathbf{C}_{\mathbf{X}}^{(i)} - \mathbf{C}_{\mathbf{X}\mathbf{Y}}^{(i)}(\mathbf{C}_{\mathbf{Y}}^{(i)})^{-1}\mathbf{C}_{\mathbf{Y}\mathbf{X}}^{(i)}) \preceq \sum_{i=1}^L \mathbf{C}_{\mathbf{X}}^{(i)} - \left(\sum_{i=1}^L \mathbf{C}_{\mathbf{X}\mathbf{Y}}^{(i)}\right)\left(\sum_{i=1}^L \mathbf{C}_{\mathbf{Y}}^{(i)}\right)^{-1}\left(\sum_{i=1}^L \mathbf{C}_{\mathbf{Y}\mathbf{X}}^{(i)}\right). \quad (75)$$

Particularly,

$$\mathbf{C}_{\mathbf{X}} - \mathbf{C}_{\mathbf{X}\mathbf{Y}}(\mathbf{C}_{\mathbf{Y}})^{-1}\mathbf{C}_{\mathbf{Y}\mathbf{X}} \preceq \mathbf{C}'_{\mathbf{X}} - \mathbf{C}'_{\mathbf{X}\mathbf{Y}}(\mathbf{C}'_{\mathbf{Y}})^{-1}\mathbf{C}'_{\mathbf{Y}\mathbf{X}} \quad (76)$$

for all

$$0 \preceq \begin{pmatrix} \mathbf{C}_{\mathbf{X}} & \mathbf{C}_{\mathbf{X}\mathbf{Y}} \\ \mathbf{C}_{\mathbf{Y}\mathbf{X}} & \mathbf{C}_{\mathbf{Y}} \end{pmatrix} := \mathbf{C} \preceq \mathbf{C}' := \begin{pmatrix} \mathbf{C}'_{\mathbf{X}} & \mathbf{C}'_{\mathbf{X}\mathbf{Y}} \\ \mathbf{C}'_{\mathbf{Y}\mathbf{X}} & \mathbf{C}'_{\mathbf{Y}} \end{pmatrix}.$$

Proof. Inequality (74) is a direct consequence of [30, Appendix]. For $\lambda_i \equiv 1/L$, (74) is seen as

$$\left(\sum_{i=1}^L \mathbf{C}_{\mathbf{X}\mathbf{Y}}^{(i)}\right)\left(\sum_{i=1}^L \mathbf{C}_{\mathbf{Y}}^{(i)}\right)^{-1}\left(\sum_{i=1}^L \mathbf{C}_{\mathbf{Y}\mathbf{X}}^{(i)}\right) \preceq \sum_{i=1}^L \mathbf{C}_{\mathbf{X}\mathbf{Y}}^{(i)}(\mathbf{C}_{\mathbf{Y}}^{(i)})^{-1}\mathbf{C}_{\mathbf{Y}\mathbf{X}}^{(i)},$$

which yields (75). Then, for $\mathbf{C}' = \mathbf{C} + \tilde{\mathbf{C}}$ with $\tilde{\mathbf{C}} \succeq 0$ (so $\tilde{\mathbf{C}}_{\mathbf{X}} - \tilde{\mathbf{C}}_{\mathbf{X}\mathbf{Y}}(\tilde{\mathbf{C}}_{\mathbf{Y}})^{-1}\tilde{\mathbf{C}}_{\mathbf{Y}\mathbf{X}} \succeq 0$ by Shur's complement), one has

$$\mathbf{C}'_{\mathbf{X}\mathbf{Y}}(\mathbf{C}'_{\mathbf{Y}})^{-1}\mathbf{C}'_{\mathbf{Y}\mathbf{X}} \preceq \mathbf{C}_{\mathbf{X}\mathbf{Y}}(\mathbf{C}_{\mathbf{Y}})^{-1}\mathbf{C}_{\mathbf{Y}\mathbf{X}} + \tilde{\mathbf{C}}_{\mathbf{X}\mathbf{Y}}(\tilde{\mathbf{C}}_{\mathbf{Y}})^{-1}\tilde{\mathbf{C}}_{\mathbf{Y}\mathbf{X}}.$$

Therefore,

$$\begin{aligned} \mathbf{C}_{\mathbf{X}} - \mathbf{C}_{\mathbf{X}\mathbf{Y}}(\mathbf{C}_{\mathbf{Y}})^{-1}\mathbf{C}_{\mathbf{Y}\mathbf{X}} &\preceq \\ \mathbf{C}_{\mathbf{X}} - \mathbf{C}_{\mathbf{X}\mathbf{Y}}(\mathbf{C}_{\mathbf{Y}})^{-1}\mathbf{C}_{\mathbf{Y}\mathbf{X}} + (\tilde{\mathbf{C}}_{\mathbf{X}} - \tilde{\mathbf{C}}_{\mathbf{X}\mathbf{Y}}(\tilde{\mathbf{C}}_{\mathbf{Y}})^{-1}\tilde{\mathbf{C}}_{\mathbf{Y}\mathbf{X}}) &= \\ \mathbf{C}' - (\mathbf{C}_{\mathbf{X}\mathbf{Y}}(\mathbf{C}_{\mathbf{Y}})^{-1}\mathbf{C}_{\mathbf{Y}\mathbf{X}} + \tilde{\mathbf{C}}_{\mathbf{X}\mathbf{Y}}(\tilde{\mathbf{C}}_{\mathbf{Y}})^{-1}\tilde{\mathbf{C}}_{\mathbf{Y}\mathbf{X}}) &\preceq \\ \mathbf{C}'_{\mathbf{X}} - \mathbf{C}'_{\mathbf{X}\mathbf{Y}}(\mathbf{C}'_{\mathbf{Y}})^{-1}\mathbf{C}'_{\mathbf{Y}\mathbf{X}}, & \end{aligned}$$

proving (76).⁷ \square

Important matrix inequalities for covariances are summarized in the following theorem.

Theorem 3: For a jointly GM characterized by equation (1), the following matrix inequalities hold true

$$\mathbf{C}_{mmse}(\mathbf{y}) \preceq \mathbf{C}_{lmse} \quad \forall \mathbf{y}, \quad (77)$$

$$\sum_{i=1}^L \lambda_i \mathbf{C}_{\mathbf{X}^{(i)}|\mathbf{Y}^{(i)}} \preceq \mathbb{E}_{\mathbf{y}}(\mathbf{C}_{mmse}(\mathbf{y})) \preceq \mathbf{C}_{lmse}, \quad (78)$$

$$\sum_{i=1}^L \lambda_i \mathbf{C}_{\mathbf{X}^{(i)}|\mathbf{Y}^{(i)}} \preceq \mathbf{C}_{\mathbf{X}|\mathbf{Y}}(\lambda) \preceq \mathbf{C}_{lmse}, \quad (79)$$

where $\mathbf{C}_{\mathbf{X}|\mathbf{Y}}(\lambda) = \mathbf{C}_{\mathbf{X}}(\lambda) - \mathbf{C}_{\mathbf{X}\mathbf{Y}}(\lambda)\mathbf{C}_{\mathbf{Y}}^{-1}(\lambda)\mathbf{C}_{\mathbf{Y}\mathbf{X}}(\lambda)$.

Proof. (77) follows directly from the definitions (63), (65) and (69):

$$\begin{aligned} \mathbf{C}_{lmse}(\mathbf{y}) - \mathbf{C}_{mmse} &= \\ \int (\mathbf{x}^2) f_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x})d\mathbf{x} - \hat{\mathbf{x}}_{lmse} \int \mathbf{x}^T f_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x})d\mathbf{x} &= \\ - \int \mathbf{x} f_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x})d\mathbf{x} + (\mathbf{x}_{lmse})^2 - \mathbf{C}_{mmse} &= \\ \mathbf{C}_{mmse} + \mathbf{x}_{mmse}^2 - \hat{\mathbf{x}}_{lmse}(\hat{\mathbf{x}}_{mmse})^T &= \\ - \hat{\mathbf{x}}_{mmse}\hat{\mathbf{x}}_{lmse}^T + (\mathbf{x}_{lmse})^2 - \mathbf{C}_{mmse} &= \\ (\hat{\mathbf{x}}_{lmse} - \hat{\mathbf{x}}_{mmse})^2 \succeq &\succeq \\ 0. & \quad (80) \end{aligned}$$

Also, by (60) $\sum_{i=1}^L \lambda_i(\mathbf{y})(\mathbf{m}_{\mathbf{X}^{(i)}|\mathbf{Y}^{(i)}})^2 \succeq (\hat{\mathbf{x}}_{mmse})^2$, so (78) is shown as follows:

$$\begin{aligned} \mathbb{E}_{\mathbf{y}}(\mathbf{C}_{mmse}(\mathbf{y})) &\succeq \int \sum_{i=1}^L \lambda_i(\mathbf{y})\mathbf{C}_{\mathbf{X}^{(i)}|\mathbf{Y}^{(i)}} f_{\mathbf{Y}}(\mathbf{y})d\mathbf{y} \\ &= \int \sum_{i=1}^L \lambda_i \mathcal{N}(\mathbf{y}; \mathbf{m}_{\mathbf{Y}}^{(i)}, \mathbf{C}_{\mathbf{Y}}^{(i)})\mathbf{C}_{\mathbf{X}^{(i)}|\mathbf{Y}^{(i)}} d\mathbf{y} \\ &= \sum_{i=1}^L \lambda_i \mathbf{C}_{\mathbf{X}^{(i)}|\mathbf{Y}^{(i)}}. \end{aligned}$$

Finally, (79) is a direct consequence of Lemma 1. \square

To our best knowledge, the matrix inequalities (77)-(79) have not been known in literature. Particularly, (78) implies the main result of [37]: $\sum_{i=1}^L \lambda_i \text{Trace}(\mathbf{C}_{\mathbf{X}^{(i)}|\mathbf{Y}^{(i)}}) \leq$

$\mathbb{E}_{\mathbf{y}}(\epsilon_{mmse}^2(\mathbf{y})) \leq \epsilon_{lmse}^2$, which was proved by many involved calculations.

⁷Shur's complement holds for $\tilde{\mathbf{C}}_{\mathbf{Y}} \succeq 0$ with $(\tilde{\mathbf{C}}_{\mathbf{Y}})^{-1}$ replaced by the pseudoinverse of $\tilde{\mathbf{C}}_{\mathbf{Y}}$, under which all the derivations are still valid.

APPENDIX B: ONE-HOP COMMUNICATION

In one-hop communication between the sensors and FC, the relay plays the role of the FC, so

$$\begin{aligned} \mathbf{C}_{\mathbf{X}|\mathbf{Z}} &= (\mathbf{C}_{\mathbf{X}} - \mathbf{C}_{\mathbf{Y}\mathbf{X}}^T(\mathbf{C}_{\mathbf{Y}})^{-1}\mathbf{C}_{\mathbf{Y}\mathbf{X}}) \\ &+ \mathbf{C}_{\mathbf{Y}\mathbf{X}}^T(\mathbf{C}_{\mathbf{Y}})^{-1}((\mathbf{C}_{\mathbf{Y}})^{-1} \\ &+ \text{diag}[\varphi_j(\alpha_j)]_1^M)^{-1}(\mathbf{C}_{\mathbf{Y}})^{-1}\mathbf{C}_{\mathbf{Y}\mathbf{X}} \end{aligned} \quad (81)$$

with $\varphi_j(\alpha_j, \beta_j) = h_{jR} \frac{\alpha_j}{\sigma_j}$. Recall that h_{jR} is the channel gain from the sensor j to the FC, and σ_{jR} is the power of the background noise at FC/relay.

Consider

$$\min_{\alpha > 0} \text{Trace}(\mathbf{C}_{lms\epsilon}(\alpha)) \quad \text{s.t.} \quad (4),$$

which is equivalent to the following program

$$\begin{aligned} \min_{\alpha > 0} \varphi(\alpha) &:= \text{Trace}(\Psi^T(\Phi + \text{diag}[\varphi_j(\alpha_j)]_1^M)^{-1}\Psi) \\ &\quad \text{s.t.} \quad (4), \end{aligned} \quad (82)$$

where $\Psi = \mathbf{C}_{\mathbf{Y}}^{-1}\mathbf{C}_{\mathbf{Y}\mathbf{X}}$, $\Phi = \mathbf{C}_{\mathbf{Y}}^{-1}$. Unlike (19), the program (82) is convex, which has been solved in [22] by semi-definite programming (SDP). The complexity of SDP is still high for online applications and more importantly, it is not scalable. We now develop a path-following scalable procedure for the computational solution of (82).

Given $\alpha^{(\kappa)}$ feasible to (4), we now process the following successive approximation. Define

$$\begin{aligned} \varphi_j^{(\kappa)} &= \varphi_j(\alpha_j^{(\kappa)}), \\ 0 < \Theta^{(\kappa)} &= \text{diag}[\varphi_j^{(\kappa)}]_1^M (\Phi + \text{diag}[\varphi_j^{(\kappa)}]_1^M)^{-1} \Psi \Psi^T \\ &\quad \times (\Phi + \text{diag}[\varphi_j^{(\kappa)}]_1^M)^{-1} \text{diag}[\varphi_j^{(\kappa)}]_1^M, \\ 0 < \rho_j^{(\kappa)} &= \Theta^{(\kappa)}(j, j), \end{aligned}$$

where $\Theta^{(\kappa)}(j, j)$ is the j -th diagonal entry of $\Theta^{(\kappa)}$. Analogously to Theorem 1 we can show that

$$\varphi(\alpha) \leq \varphi^{(\kappa)}(\alpha) := \varphi(\alpha^{(\kappa)}, \beta^{(\kappa)}) + \sum_{j=1}^M \rho_j^{(\kappa)} \left(\frac{\sigma_{jR}}{h_{jR}\alpha_j} - \frac{1}{\varphi_j^{(\kappa)}} \right).$$

Accordingly, we consider the majorant minimization

$$\min_{\alpha} \varphi^{(\kappa)}(\alpha) \quad \text{s.t.} \quad (4),$$

which admits the optimal solution in closed-form

$$\alpha_j^{(\kappa+1)} = \lambda_T \sqrt{\rho_j^{(\kappa)} \sigma_{jR} / h_{jR}} \quad (83)$$

where $\lambda_T > 0$ such that $\alpha_j^{(\kappa+1)}$ satisfies the power constraint (4), i.e.

$$\lambda_T = P_T / \sum_{j=1}^M \mathbb{E}[|y_j|^2] \sqrt{\rho_j^{(\kappa)} \sigma_{jR} / h_{jR}}.$$

Algorithm 2 is a pseudocode for solving (82). The limit point by Algorithm 2 is the global optimal solution of (82) because it satisfies KKT conditions of the convex program (82).

Algorithm 2 Fast iterative procedure for 1H power allocation

- 1: Initialize $\kappa := 0$ and $\alpha^{(0)}$ feasible to (4).
- 2: **repeat** Generate a feasible solution $\alpha^{(\kappa+1)}$ according to formula (83) .
- 3: **until**
$$\frac{\varphi(\alpha^{(\kappa)}) - \varphi(\alpha^{(\kappa+1)})}{\varphi(\alpha^{(\kappa)})} \leq \epsilon \quad (84)$$
 for a given tolerance ϵ .
- 4: Extract $\alpha^* = \alpha^{(\kappa)}$ as the solution of 1H power allocation.

References

- [1] I. Akyildiz, W. Su, Y. Sankarsubramaniam, and E. Cayirci, "Wireless sensor networks : A survey," *Comput. Netw.*, vol. 38, pp. 393–422, 2002.
- [2] D. Nguyen and M. Bagajewicz, "Design of nonlinear sensor networks for process plants," *Industrial & Engineering Chemistry Research*, vol. 47, no. 15, pp. 5529–5542, 2008.
- [3] S. Kim, B. Ku, W. Hong, and H. Ko, "Performance comparison of target localization for active sonar systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 4, pp. 1371–1380, 2008.
- [4] K. Ho and L. M. Vicente, "Sensor allocation for source localization with decoupled range and bearing estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 5773–5789, 2008.
- [5] X. Zhang, H. V. Poor, and M. Chiang, "Optimal power allocation for distributed detection over MIMO channels in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4124–4141, 2008.
- [6] J. Majchrzak, M. Michalski, and G. Wiczynski, "Distance estimation with a long-range ultrasonic sensor system," *IEEE Sens. J.*, vol. 9, no. 7, pp. 767–773, 2009.
- [7] S. Kar and J. Moura, "A mixed time-scale algorithm for distributed parameter estimation: Nonlinear observation models and imperfect communication," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP) 2009*, 2009, pp. 3669–3672.
- [8] M. Arik and O. Akan, "Collaborative mobile target imaging in UWB wireless radar sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 6, pp. 950–961, 2010.
- [9] L. Liu, X. Zhang, and H. Ma, "Optimal node selection for target localization in wireless camera sensor networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 7, pp. 3562–3576, 2010.
- [10] —, "Percolation theory-based explosure-apth prevention for wireless sensor networks coverage in internet of things," *IEEE Sensor J.*, vol. 13, no. 10, pp. 3625–3636, 2013.
- [11] Y. Zhou, C. Huang, T. Jang, and S. Cui, "Wireless sensor networks and internet of things: optimal estimation with nonuniform quantization and bandwidth allocation," *IEEE Sensor J.*, vol. 13, no. 10, pp. 3568–3574, 2013.
- [12] G. Alirezaei, M. Reyer, and R. Mathar, "Optimum power allocation in sensor networks for passive radar applications," *IEEE Trans. Wireless Commun.*, vol. 13, no. 6, pp. 3222–3231, 2014.
- [13] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," *Lecture Notes in Computer Science*, vol. 2634, pp. 162–177, 2003.
- [14] —, "Power, spatio-temporal bandwidth, and distortion in large sensor network," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 745–754, 2005.
- [15] H. V. Poor, *An Introduction to Signal Detection and Estimation (second edition)*. New York: Springer-Verlag, 1994.
- [16] S. Cui, J. Xiao, A. Goldsmith, Z. Luo, and H. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4683–4695, 2007.
- [17] G. Thattai and U. Mitra, "Sensor selection and power allocation for distributed estimation in sensor networks: Beyond the star topology," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2649–2661, 2008.
- [18] I. Bahceci and A. Khandani, "Linear estimation of correlated data in wireless sensor networks with optimum power allocation and analog modulation," *IEEE Trans. Commun.*, vol. 56, no. 7, pp. 1146–1156, 2008.
- [19] J. Fang and H. Li, "Power constrained distributed estimation with correlated sensor data," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3292–3297, 2009.

- [20] M. H. Chaudhary and L. Vandendorpe, "Power constrained linear estimation in wireless sensor networks with correlated data and digital modulation," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 570–584, 2012.
- [21] —, "Performance of power -constrained estimation in hierarchical wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 724–739, 2013.
- [22] U. Rashid, H. D. Tuan, P. Apkarian, and H. H. Kha, "Globally optimized power allocation in multiple sensor fusion for linear and nonlinear networks," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 903–915, 2012.
- [23] R. Pabst et al, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Commun. Magazine*, vol. 42, no. 9, pp. 80–89, 2004.
- [24] G. J. McLachlan and K. E. Basford, *Mixture models: inference and applications to clustering*. New York: Marcel Dekker, 1988.
- [25] D. Persson, T. Eriksson, and P. Hedelin, "Packet video error concealment with Gaussian mixture models," *IEEE Trans. Image Processing*, vol. 17, no. 2, pp. 145–154, 2008.
- [26] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From Gaussian mixture models to structured sparsity," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2481–2499, 2012.
- [27] J. Flam, D. Zahariah, M. Vehkaperä, and S. Chatterjee, "The linear model under mixed Gaussian inputs: designing the transfer matrix," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5247–5259, 2013.
- [28] S. A. Pasha, H. D. Tuan, and B. Vo, "Nonlinear Bayesian filtering using the unscented linear fractional transformation model," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 477–489, 2010.
- [29] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
- [30] U. Rashid, H. D. Tuan, H. H. Kha, and H. H. Nguyen, "Joint optimization of source precoding and relay beamforming in wireless MIMO relay networks," *IEEE Trans. Commun.*, vol. 62, no. 2, pp. 488–499, 2014.
- [31] H. Tuy, *Convex Analysis and Global Optimization (second edition)*. Springer, 2016.
- [32] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Operations Research*, vol. 26, no. 4, pp. 681–683, Jul. 1978.
- [33] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge University Press, 2005.
- [34] S. Julier, J. Uhlmann, and H. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Trans. Autom. Control*, vol. 45, no. 3, pp. 477–482, 2000.
- [35] Y. Bar-Shalom, X.-R. Li, and T. Kirubajan, *Multi-target-multi-sensor tracking: principles and techniques*. New York: Wiley, 2001.
- [36] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice Hall, 1993.
- [37] J. Flam, S. Chatterjee, K. Kansanen, and T. Ekman, "On MMSE estimation: A linear model under Gaussian mixture statistics," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3840–3845, 2012.



Johann A. Bengua received the B.E. degree in telecommunications engineering from the University of New South Wales, Kensington, Australia, in 2012 and Ph.D. degree in electrical engineering from the University of Technology Sydney, Ultimo, Australia, in 2017. He is now a data scientist for Teradata Australia and New Zealand. His current research interests include tensor networks, machine learning, image and video processing.



Hoang Duong Tuan received the Diploma (Hons.) and Ph.D. degrees in applied mathematics from Odessa State University, Ukraine, in 1987 and 1991, respectively. He spent nine academic years in Japan as an Assistant Professor in the Department of Electronic-Mechanical Engineering, Nagoya University, from 1994 to 1999, and then as an Associate Professor in the Department of Electrical and Computer Engineering, Toyota Technological Institute, Nagoya, from 1999 to 2003. He was a Professor with the School of Electrical Engineering and Telecommunications, University of New South Wales, from 2003 to 2011. He is currently a Professor with the School of Electrical and Data Engineering, University of Technology Sydney. He has been involved in research with the areas of optimization, control, signal processing, wireless communication, and biomedical engineering for more than 20 years.



Trung Q. Duong (S'05, M'12, SM'13) received his Ph.D. degree in Telecommunications Systems from Blekinge Institute of Technology (BTH), Sweden in 2012. Since 2013, he has joined Queen's University Belfast, UK as a Lecturer (Assistant Professor). His current research interests include small-cell networks, physical layer security, energy-harvesting communications, cognitive relay networks. He is the author or co-author of more than 260 technical papers published in scientific journals (142 articles) and presented at international conferences (121 papers).

Dr. Duong currently serves as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON COMMUNICATIONS, IET COMMUNICATIONS, and a Senior Editor for IEEE COMMUNICATIONS LETTERS. He was awarded the Best Paper Award at the IEEE Vehicular Technology Conference (VTC-Spring) in 2013, IEEE International Conference on Communications (ICC) 2014, and IEEE Global Communications Conference (GLOBECOM) 2016. He is the recipient of prestigious Royal Academy of Engineering Research Fellowship (2016-2021).



H. Vincent Poor (S72, M77, SM82, F87) received the Ph.D. degree in EECS from Princeton University in 1977. From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990 he has been on the faculty at Princeton, where he is the Michael Henry Strater University Professor of Electrical Engineering. From 2006 till 2016, he served as Dean of Princeton's School of Engineering and Applied Science. He has also held visiting appointments at several other institutions, most recently at Berkeley and Cambridge. His research interests are in the areas of information theory and signal processing, and their applications in wireless networks and related fields. Among his publications in these areas is the recent book *Information Theoretic Security and Privacy of Information Systems* (Cambridge University Press, 2017).

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, and a foreign member of the Royal Society. In 1990, he served as President of the IEEE Information Theory Society, and in 2004-07 as the Editor-in-Chief of the IEEE TRANSACTIONS ON INFORMATION THEORY. He received the Technical Achievement and Society Awards of the IEEE Signal Processing Society in 2007 and 2011, respectively. Recent recognition of his work includes the 2017 IEEE Alexander Graham Bell Medal, a D.Sc. *honoris causa* from Syracuse University awarded in 2017, and election as a Foreign Member of the National Academy of Engineering of Korea and an Honorary Member of the National Academy of Sciences of Korea, both in 2017.